

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 8, 2003, 19:05:52 ; Search time 4068 Seconds
(without alignments)
10444.960 Million cell updates/sec

Title: US-09-595-947C-1
Perfect score: 1460
Sequence: 1 gcaggtacgagagagcag.....agagtgaactaatccagtgt 1460

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 2034640 seqs, 14551402878 residues

Word size : 15

Total number of hits satisfying chosen parameters: 42842

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_hgt.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vl.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vl.*
- 30: em_hgt_hum.*
- 31: em_hgt_inv.*
- 32: em_hgt_other.*
- 33: em_hgt_mus.*
- 34: em_hgt_pln.*
- 35: em_hgt_rod.*
- 36: em_hgt_nam.*
- 37: em_hgt_vrt.*
- 38: em_sy.*
- 39: em_hgtgo_hum.*
- 40: em_hgtgo_mus.*
- 41: em_hgtgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1460	100.0	1491	6	A91167 Sequence 1
2	1460	100.0	1491	10	Y10619 R.norvegicus
3	634	43.4	166677	2	AC127817 Rattus no
4	92	6.3	861	10	U76208 Mus musculus
5	92	6.3	1861	10	AF364300 Mus muscu
6	92	6.3	5567	10	Y0167 M.musculus
7	92	6.3	138070	2	AC109783 Mus muscu
8	92	6.3	185806	2	AC127417 Mus muscu
9	35	2.4	6123	9	AF303002 Homo sapi
10	35	2.4	91531	2	AC079846_3 Continuation (4 of
11	35	2.4	179697	9	AC023886 Homo sapi
12	32	2.2	1330	9	AL133776 Homo sapi
13	32	2.2	5340	9	AF234829 Homo sapi
14	32	2.2	165110	9	AL450311 Human DNA
15	32	2.2	173341	2	AC021954 Homo sapi
16	26	1.8	170896	2	AC011010 Homo sapi
17	25	1.7	25	6	A91170 Sequence 4
18	23	1.6	735	10	U67776 Mus musculus
19	23	1.6	770	5	AF123884 Gallus ga
20	23	1.6	790	5	GA012659 Gallus ga
21	23	1.6	932	10	MMATH4C Y09166 M.musculus
22	23	1.6	1315	10	MMU63841 Mus muscu
23	23	1.6	1333	6	AR023715 Sequence
24	23	1.6	1341	5	AF109014 Gallus ga
25	23	1.6	1385	10	MMU76207 U76207 Mus musculus
26	23	1.6	1412	10	MMATH4A Y07621 M.musculus
27	23	1.6	1880	5	AF303000 Gallus ga
28	23	1.6	10393	10	AF303001 Mus muscu
29	23	1.6	71538	2	AC118243 Mus muscu
30	23	1.6	123855	2	AC102600 Mus muscu
31	23	1.6	174688	2	AC124395 Mus muscu
32	21	1.4	310	6	I84656 Sequence 12
33	21	1.4	1164	5	GGNEURODL Y09596 G.gallus mr
34	21	1.4	1268	6	AR023709 Sequence
35	21	1.4	1268	9	HSU63842 Human neuro
36	21	1.4	1352	6	I84655 Sequence 10
37	21	1.4	1535	6	AR023708 Sequence
38	21	1.4	1550	6	AR103242 Sequence
39	21	1.4	1550	6	BD000105 Different
40	21	1.4	1550	9	HSU58681 Homo sapien
41	21	1.4	1596	5	AF060885 Gallus ga
42	21	1.4	1675	9	BC008687 Homo sapi
43	21	1.4	1717	9	BC028226 Homo sapi
44	21	1.4	1831	10	MUSNDRF D83507 Mouse mRNa
45	21	1.4	1843	10	RAT4 D82868 Rattus norv

ALIGNMENTS

RESULT 1	A91167	Sequence 1	1491 bp	DNA	linear	PAT 22-JAN-2000
LOCUS	A91167	Sequence 1 from Patent WO9827206.				
DEFINITION	A91167					
ACCESSION	A91167					
VERSION	A91167.1	GI:6740202				
KEYWORDS						
SOURCE	Rattus sp.					
ORGANISM	Rattus sp.					
	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;					
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;					
	Rattus.					
REFERENCE	1 (bases 1 to 1491)					
AUTHORS	Icard-Liepkalns,C., Mallet,J. and Corresponding,N.A.					
JOURNAL	Patent: WO 9827206-A 1 25-JUN-1998;					

ICARD LIEPKALNS CHRISTINE (FR); MALLET JACQUES (FR)

FEATURES

Location/Qualifiers

1. 1491

/organism="Rattus sp."

/db_xref="taxon:10118"

459..1103

/note="unnamed protein product"

/codon_start=1

/protein_id="CAP69414.1"

/db_xref="GI:6740203"

/translation="MAPHPLDAPTIOVSOETQOPFGASDHVLSNTPSPPTLVPR

DSEARAGCRGTGSKLRARRGRNRPKSELALSKORRRRRKANDRENRHNLNSA

LUALRVLTFFDDAKLTIELTFAHNTIWAOTFLRIADHSFYGPPEPVPCEIGLS

PGGSGDNGWSIYSPVQAGSLSPTRASLEFFPLQVPPSPCLLPGLTVFSDFL"

BASE COUNT 307 a 487 c 413 g 284 t

ORIGIN

Query Match 100.0%; Score 1460; DB 6; Length 1491;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1460; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAGGTAGCGAGAGGAGCAGTCCCTGGGCCCGCCGTTGCTGATTGGCCCGTGGCAGACGGCA 60

DB 1 GCAGGTAGCGAGAGGAGCAGTCCCTGGGCCCGCCGTTGCTGATTGGCCCGTGGCAGACGGCA 60

QY 61 GCAGCCGGCAGGACGCTCCTGGTCCGGCGAGAGCAGATAAGCGTGCAGGGGACACA 120

DB 61 GCAGCCGGCAGGACGCTCCTGGTCCGGCGAGAGCAGATAAGCGTGCAGGGGACACA 120

QY 121 CGATTAGCAGCTCAGAGTCCCTGCTGGTCTCACCAGTGCACAGAGCCGAGACCCCT 180

DB 121 CGATTAGCAGCTCAGAGTCCCTGCTGGTCTCACCAGTGCACAGAGCCGAGACCCCT 180

QY 181 CCGAGCTTCTTTGCTGCTCCAGACGCAATTTACTCCAGGCGAGGGCGCTTGACGCTCAG 240

DB 181 CCGAGCTTCTTTGCTGCTCCAGACGCAATTTACTCCAGGCGAGGGCGCTTGACGCTCAG 240

QY 241 CAACACTTCGAGGAGCAGAGGGTTACGATATCCACCGCTGCTGACGCTGACGACCC 300

DB 241 CAACACTTCGAGGAGCAGAGGGTTACGATATCCACCGCTGCTGACGCTGACGACCC 300

QY 301 GCAGCTCTGTCTTTGAGCCGGAGTAACTAGGTAACATTTAGGAACCTCCAAAGGG 360

DB 301 GCAGCTCTGTCTTTGAGCCGGAGTAACTAGGTAACATTTAGGAACCTCCAAAGGG 360

QY 361 TAGAAGAGGAGTGGTGGCGTACTCTAGTCCCGGCTGGAGTGACCTCTAAGTCAGAG 420

DB 361 TAGAAGAGGAGTGGTGGCGTACTCTAGTCCCGGCTGGAGTGACCTCTAAGTCAGAG 420

QY 421 ACTGTACACCCCTTCCATTTTCCCAACCTCAGAGTGGCGCTCATCCCTTGGATG 480

DB 421 ACTGTACACCCCTTCCATTTTCCCAACCTCAGAGTGGCGCTCATCCCTTGGATG 480

QY 481 CGCCACCATCCAAAGTGTCCCAAGAGACCCAGCAACCCCTTCCGCGAGCCTCGGACACG 540

DB 481 CGCCACCATCCAAAGTGTCCCAAGAGACCCAGCAACCCCTTCCGCGAGCCTCGGACACG 540

QY 541 AAGTGCTCAGTTCOAATTCACCCACCTAGCCCTACTCTCGTACCGAGGAGTGTCCG 600

DB 541 AAGTGCTCAGTTCOAATTCACCCACCTAGCCCTACTCTCGTACCGAGGAGTGTCCG 600

QY 601 AAGCAGAGCAGTGTACTCGGAGGACATCGAGGAGTCCGTCGCGCGCGGAGGGC 660

DB 601 AAGCAGAGCAGTGTACTCGGAGGACATCGAGGAGTCCGTCGCGCGCGGAGGGC 660

QY 661 GCACAGGCCCAAGAGCGAGTGTGCTAGCAGCAGCAGCAGGAGCGCGCAAGAGG 720

DB 661 GCACAGGCCCAAGAGCGAGTGTGCTAGCAGCAGCAGCAGGAGCGCGCAAGAGG 720

QY 721 CCAACGACCGGAGCGCAACCGCATGACACACCTTAACCTCGGCGCTGGATCGCTGCGG 780

DB 721 CCAACGACCGGAGCGCAACCGCATGACACACCTTAACCTCGGCGCTGGATCGCTGCGG 780

QY 781 GTGTCTGCCACCTTCCCGGATGACGCCAAACTTACAAAGATCGAGACCTTGGGCTTCG 840
 DB 781 GTGTCTGCCACCTTCCCGGATGACGCCAAACTTACAAAGATCGAGACCTTGGGCTTCG 840

QY 841 CCACAACTACATTTGGGCACTGACTCAGAGCGTGGCGATAGCGGACACAGCTTCTACG 900
 DB 841 CCACAACTACATTTGGGCACTGACTCAGAGCGTGGCGATAGCGGACACAGCTTCTACG 900

QY 901 GCCCGAGCCCTGTGCCCTGTGGGAGCTGGGAACCCCGGAGGGGTCCACAGCGCG 960
 DB 901 GCCCGAGCCCTGTGCCCTGTGGGAGCTGGGAACCCCGGAGGGGTCCACAGCGCG 960

QY 961 ACTGGGGCTATCTACTCTCCAGTTTCCCAAGCTGTTAGCCTGAGCCCAACAGCTCAT 1020
 DB 961 ACTGGGGCTATCTACTCTCCAGTTTCCCAAGCTGTTAGCCTGAGCCCAACAGCTCAT 1020

QY 1021 TGGAGGAGTTCCCTGGCTGCAGTGCAGTCCAGCTCCCATCTCTGCTCTCCGGGACCC 1080
 DB 1021 TGGAGGAGTTCCCTGGCTGCAGTGCAGTCCAGCTCCCATCTCTGCTCTCCGGGACCC 1080

QY 1081 TGTGTCTCTCAGACTTCTTGTGAAGGCCCCAAACAGAGCCCTGGGCGTGGCGCTGCGCAG 1140
 DB 1081 TGTGTCTCTCAGACTTCTTGTGAAGGCCCCAAACAGAGCCCTGGGCGTGGCGCTGCGCAG 1140

QY 1141 AAAGGAGGAGTGCAGAGCTGTCTGAAATGGAAGGTAGTGGAGCAGCTCGAGCATCTCGC 1200
 DB 1141 AAAGGAGGAGTGCAGAGCTGTCTGAAATGGAAGGTAGTGGAGCAGCTCGAGCATCTCGC 1200

QY 1201 CCTTCTGGCTTTCATTAGTCAGTCCCTGATTAAACAGGATTTCGCACAGTTCCTTGTCT 1260
 DB 1201 CCTTCTGGCTTTCATTAGTCAGTCCCTGATTAAACAGGATTTCGCACAGTTCCTTGTCT 1260

QY 1261 GCTGTGCTGCACAAAGGACATTCAGGCTGATCTCTTAAACCTCCTCAGTGTGGCC 1320
 DB 1261 GCTGTGCTGCACAAAGGACATTCAGGCTGATCTCTTAAACCTCCTCAGTGTGGCC 1320

QY 1321 ACCTCAACTCCCGCTCCAGCAGAGGAGAGCCGTACACTAATAGTTGGGAGCTCC 1380
 DB 1321 ACCTCAACTCCCGCTCCAGCAGAGGAGAGCCGTACACTAATAGTTGGGAGCTCC 1380

QY 1381 ATACTTCTGCTGACTCCGCCCTCTTTCAAATCTGCGGGCTCCCAACACCGCTTCTCC 1440
 DB 1381 ATACTTCTGCTGACTCCGCCCTCTTTCAAATCTGCGGGCTCCCAACACCGCTTCTCC 1440

QY 1441 AGAGTGACCTTAATCCAGTGT 1460
 DB 1441 AGAGTGACCTTAATCCAGTGT 1460

RESULT 2

RNRELAXT

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

RNRELAXT 1491 bp mRNA linear ROD 06-MAY-1997
 R.norvegicus mRNA for transcriptional regulator, Relax.

Y10619
 Y10619.1 GI:2072737

Relax; transcriptional regulator.

Rattus norvegicus.

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

1 (bases 1 to 1491)

Ravassard, P., Chatail, F., Mallet, J. and Icard-Liepkalns, C.

Relax, a novel rat bHLH transcriptional regulator transiently

expressed in the ventricular proliferating zone of the developing

central nervous system

J. Neurosci. Res. 48 (2), 146-158 (1997)

97276390

9130143

2 (bases 1 to 1491)

Ravassard, P.

Direct Submission

Submitted (20-JAN-1997) P. Ravassard, CNRS UMR 9923, Bat. CERVI,

Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,
 Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,
 Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,
 Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,
 Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,
 Douthwaite, R.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
 Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,
 Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,
 Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
 Correll, J.H., Guevara, M., Gunaratne, P., Hale, S., Hamilton, K.,
 Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,
 Hernandez, O., Hodgson, A., Hogue, M., Holloway, C., Hollins, B.,
 Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E.,
 Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,
 Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,
 Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
 Li, J., Li, Z., Lichtarge, O., Lied, C., Liu, J., Liu, W., Louised, H.,
 Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
 Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,
 Massey, E., Mawhiney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M.,
 Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
 Moser, M., Neal, D., Netsch, J., Newton, N., Nguyen, A., Nguyen, N.,
 Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogih, M., Okwuonu, G.,
 Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
 Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,
 Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savary, G.,
 Scherer, S., Scott, G., Shen, H., Shoohtari, N., Sisson, I.,
 Sodergren, E., Sonalike, T., Sparks, A., Stanley, H., Stone, H.,
 Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
 Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
 Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q.,
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
 Williams, G., Williamson, A., Wleczky, R., Wooden, S., Worley, K.,
 Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
 Weinstock, G., and Gibbs, R.
 Direct Submission
 Unpublished
 2 (bases 1 to 166677)
 Worley, K.C.
 Direct Submission
 Submitted (19-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 ----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GZXS
 Center clone name: CH230-259c16
 ----- Summary Statistics
 Sequencing vector: Plasmid;
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 73194 bases at least Q40
 Consensus quality: 79767 bases at least Q30
 Consensus quality: 84278 bases at least Q20

 * NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 consists of 80 contigs. The true order of the pieces
 is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 as soon as it is available and the accession number will
 be preserved.
 * 1 1032: contig of 1032 bp in length
 * 1033 1132: gap of unknown length
 * 1133 2146: contig of 1014 bp in length
 * 2147 2246: gap of unknown length
 * 2247 2247: contig of 1014 bp in length
 * 2248 2248: gap of unknown length
 * 2249 2249: contig of 1014 bp in length
 * 2250 2250: gap of unknown length
 * 2251 2251: contig of 1014 bp in length
 * 2252 2252: gap of unknown length
 * 2253 2253: contig of 1014 bp in length
 * 2254 2254: gap of unknown length
 * 2255 2255: contig of 1014 bp in length
 * 2256 2256: gap of unknown length
 * 2257 2257: contig of 1014 bp in length
 * 2258 2258: gap of unknown length
 * 2259 2259: contig of 1014 bp in length
 * 2260 2260: gap of unknown length
 * 2261 2261: contig of 1014 bp in length
 * 2262 2262: gap of unknown length
 * 2263 2263: contig of 1014 bp in length
 * 2264 2264: gap of unknown length
 * 2265 2265: contig of 1014 bp in length
 * 2266 2266: gap of unknown length
 * 2267 2267: contig of 1014 bp in length
 * 2268 2268: gap of unknown length
 * 2269 2269: contig of 1014 bp in length
 * 2270 2270: gap of unknown length
 * 2271 2271: contig of 1014 bp in length
 * 2272 2272: gap of unknown length
 * 2273 2273: contig of 1014 bp in length
 * 2274 2274: gap of unknown length
 * 2275 2275: contig of 1014 bp in length
 * 2276 2276: gap of unknown length
 * 2277 2277: contig of 1014 bp in length
 * 2278 2278: gap of unknown length
 * 2279 2279: contig of 1014 bp in length
 * 2280 2280: gap of unknown length
 * 2281 2281: contig of 1014 bp in length
 * 2282 2282: gap of unknown length
 * 2283 2283: contig of 1014 bp in length
 * 2284 2284: gap of unknown length
 * 2285 2285: contig of 1014 bp in length
 * 2286 2286: gap of unknown length
 * 2287 2287: contig of 1014 bp in length
 * 2288 2288: gap of unknown length
 * 2289 2289: contig of 1014 bp in length
 * 2290 2290: gap of unknown length
 * 2291 2291: contig of 1014 bp in length
 * 2292 2292: gap of unknown length
 * 2293 2293: contig of 1014 bp in length
 * 2294 2294: gap of unknown length
 * 2295 2295: contig of 1014 bp in length
 * 2296 2296: gap of unknown length
 * 2297 2297: contig of 1014 bp in length
 * 2298 2298: gap of unknown length
 * 2299 2299: contig of 1014 bp in length
 * 2300 2300: gap of unknown length
 * 2301 2301: contig of 1014 bp in length
 * 2302 2302: gap of unknown length
 * 2303 2303: contig of 1014 bp in length
 * 2304 2304: gap of unknown length
 * 2305 2305: contig of 1014 bp in length
 * 2306 2306: gap of unknown length
 * 2307 2307: contig of 1014 bp in length
 * 2308 2308: gap of unknown length
 * 2309 2309: contig of 1014 bp in length
 * 2310 2310: gap of unknown length
 * 2311 2311: contig of 1014 bp in length
 * 2312 2312: gap of unknown length
 * 2313 2313: contig of 1014 bp in length
 * 2314 2314: gap of unknown length
 * 2315 2315: contig of 1014 bp in length
 * 2316 2316: gap of unknown length
 * 2317 2317: contig of 1014 bp in length
 * 2318 2318: gap of unknown length
 * 2319 2319: contig of 1014 bp in length
 * 2320 2320: gap of unknown length
 * 2321 2321: contig of 1014 bp in length
 * 2322 2322: gap of unknown length
 * 2323 2323: contig of 1014 bp in length
 * 2324 2324: gap of unknown length
 * 2325 2325: contig of 1014 bp in length
 * 2326 2326: gap of unknown length
 * 2327 2327: contig of 1014 bp in length
 * 2328 2328: gap of unknown length
 * 2329 2329: contig of 1014 bp in length
 * 2330 2330: gap of unknown length
 * 2331 2331: contig of 1014 bp in length
 * 2332 2332: gap of unknown length
 * 2333 2333: contig of 1014 bp in length
 * 2334 2334: gap of unknown length
 * 2335 2335: contig of 1014 bp in length
 * 2336 2336: gap of unknown length
 * 2337 2337: contig of 1014 bp in length
 * 2338 2338: gap of unknown length
 * 2339 2339: contig of 1014 bp in length
 * 2340 2340: gap of unknown length
 * 2341 2341: contig of 1014 bp in length
 * 2342 2342: gap of unknown length
 * 2343 2343: contig of 1014 bp in length
 * 2344 2344: gap of unknown length
 * 2345 2345: contig of 1014 bp in length
 * 2346 2346: gap of unknown length
 * 2347 2347: contig of 1014 bp in length
 * 2348 2348: gap of unknown length
 * 2349 2349: contig of 1014 bp in length
 * 2350 2350: gap of unknown length
 * 2351 2351: contig of 1014 bp in length
 * 2352 2352: gap of unknown length
 * 2353 2353: contig of 1014 bp in length
 * 2354 2354: gap of unknown length
 * 2355 2355: contig of 1014 bp in length
 * 2356 2356: gap of unknown length
 * 2357 2357: contig of 1014 bp in length
 * 2358 2358: gap of unknown length
 * 2359 2359: contig of 1014 bp in length
 * 2360 2360: gap of unknown length
 * 2361 2361: contig of 1014 bp in length
 * 2362 2362: gap of unknown length
 * 2363 2363: contig of 1014 bp in length
 * 2364 2364: gap of unknown length
 * 2365 2365: contig of 1014 bp in length
 * 2366 2366: gap of unknown length
 * 2367 2367: contig of 1014 bp in length
 * 2368 2368: gap of unknown length
 * 2369 2369: contig of 1014 bp in length
 * 2370 2370: gap of unknown length
 * 2371 2371: contig of 1014 bp in length
 * 2372 2372: gap of unknown length
 * 2373 2373: contig of 1014 bp in length
 * 2374 2374: gap of unknown length
 * 2375 2375: contig of 1014 bp in length
 * 2376 2376: gap of unknown length
 * 2377 2377: contig of 1014 bp in length
 * 2378 2378: gap of unknown length
 * 2379 2379: contig of 1014 bp in length
 * 2380 2380: gap of unknown length
 * 2381 2381: contig of 1014 bp in length
 * 2382 2382: gap of unknown length
 * 2383 2383: contig of 1014 bp in length
 * 2384 2384: gap of unknown length
 * 2385 2385: contig of 1014 bp in length
 * 2386 2386: gap of unknown length
 * 2387 2387: contig of 1014 bp in length
 * 2388 2388: gap of unknown length
 * 2389 2389: contig of 1014 bp in length
 * 2390 2390: gap of unknown length
 * 2391 2391: contig of 1014 bp in length
 * 2392 2392: gap of unknown length
 * 2393 2393: contig of 1014 bp in length
 * 2394 2394: gap of unknown length
 * 2395 2395: contig of 1014 bp in length
 * 2396 2396: gap of unknown length
 * 2397 2397: contig of 1014 bp in length
 * 2398 2398: gap of unknown length
 * 2399 2399: contig of 1014 bp in length
 * 2400 2400: gap of unknown length
 * 2401 2401: contig of 1014 bp in length
 * 2402 2402: gap of unknown length
 * 2403 2403: contig of 1014 bp in length
 * 2404 2404: gap of unknown length
 * 2405 2405: contig of 1014 bp in length
 * 2406 2406: gap of unknown length
 * 2407 2407: contig of 1014 bp in length
 * 2408 2408: gap of unknown length
 * 2409 2409: contig of 1014 bp in length
 * 2410 2410: gap of unknown length
 * 2411 2411: contig of 1014 bp in length
 * 2412 2412: gap of unknown length
 * 2413 2413: contig of 1014 bp in length
 * 2414 2414: gap of unknown length
 * 2415 2415: contig of 1014 bp in length
 * 2416 2416: gap of unknown length
 * 2417 2417: contig of 1014 bp in length
 * 2418 2418: gap of unknown length
 * 2419 2419: contig of 1014 bp in length
 * 2420 2420: gap of unknown length
 * 2421 2421: contig of 1014 bp in length
 * 2422 2422: gap of unknown length
 * 2423 2423: contig of 1014 bp in length
 * 2424 2424: gap of unknown length
 * 2425 2425: contig of 1014 bp in length
 * 2426 2426: gap of unknown length
 * 2427 2427: contig of 1014 bp in length
 * 2428 2428: gap of unknown length
 * 2429 2429: contig of 1014 bp in length
 * 2430 2430: gap of unknown length
 * 2431 2431: contig of 1014 bp in length
 * 2432 2432: gap of unknown length
 * 2433 2433: contig of 1014 bp in length
 * 2434 2434: gap of unknown length
 * 2435 2435: contig of 1014 bp in length
 * 2436 2436: gap of unknown length
 * 2437 2437: contig of 1014 bp in length
 * 2438 2438: gap of unknown length
 * 2439 2439: contig of 1014 bp in length
 * 2440 2440: gap of unknown length
 * 2441 2441: contig of 1014 bp in length
 * 2442 2442: gap of unknown length
 * 2443 2443: contig of 1014 bp in length
 * 2444 2444: gap of unknown length
 * 2445 2445: contig of 1014 bp in length
 * 2446 2446: gap of unknown length
 * 2447 2447: contig of 1014 bp in length
 * 2448 2448: gap of unknown length
 * 2449 2449: contig of 1014 bp in length
 * 2450 2450: gap of unknown length
 * 2451 2451: contig of 1014 bp in length
 * 2452 2452: gap of unknown length
 * 2453 2453: contig of 1014 bp in length
 * 2454 2454: gap of unknown length
 * 2455 2455: contig of 1014 bp in length
 * 2456 2456: gap of unknown length
 * 2457 2457: contig of 1014 bp in length
 * 2458 2458: gap of unknown length
 * 2459 2459: contig of 1014 bp in length
 * 2460 2460: gap of unknown length
 * 2461 2461: contig of 1014 bp in length
 * 2462 2462: gap of unknown length
 * 2463 2463: contig of 1014 bp in length
 * 2464 2464: gap of unknown length
 * 2465 2465: contig of 1014 bp in length
 * 2466 2466: gap of unknown length
 * 2467 2467: contig of 1014 bp in length
 * 2468 2468: gap of unknown length
 * 2469 2469: contig of 1014 bp in length
 * 2470 2470: gap of unknown length
 * 2471 2471: contig of 1014 bp in length
 * 2472 2472: gap of unknown length
 * 2473 2473: contig of 1014 bp in length
 * 2474 2474: gap of unknown length
 * 2475 2475: contig of 1014 bp in length
 * 2476 2476: gap of unknown length
 * 2477 2477: contig of 1014 bp in length
 * 2478 2478: gap of unknown length
 * 2479 2479: contig of 1014 bp in length
 * 2480 2480: gap of unknown length
 * 2481 2481: contig of 1014 bp in length
 * 2482 2482: gap of unknown length
 * 2483 2483: contig of 1014 bp in length
 * 2484 2484: gap of unknown length
 * 2485 2485: contig of 1014 bp in length
 * 2486 2486: gap of unknown length
 * 2487 2487: contig of 1014 bp in length
 * 2488 2488: gap of unknown length
 * 2489 2489: contig of 1014 bp in length
 * 2490 2490: gap of unknown length
 * 2491 2491: contig of 1014 bp in length
 * 2492 2492: gap of unknown length
 * 2493 2493: contig of 1014 bp in length
 * 2494 2494: gap of unknown length
 * 2495 2495: contig of 1014 bp in length
 * 2496 2496: gap of unknown length
 * 2497 2497: contig of 1014 bp in length
 * 2498 2498: gap of unknown length
 * 2499 2499: contig of 1014 bp in length
 * 2500 2500: gap of unknown length
 * 2501 2501: contig of 1014 bp in length
 * 2502 2502: gap of unknown length
 * 2503 2503: contig of 1014 bp in length
 * 2504 2504: gap of unknown length
 * 2505 2505: contig of 1014 bp in length
 * 2506 2506: gap of unknown length
 * 2507 2507: contig of 1014 bp in length
 * 2508 2508: gap of unknown length
 * 2509 2509: contig of 1014 bp in length
 * 2510 2510: gap of unknown length
 * 2511 2511: contig of 1014 bp in length
 * 2512 2512: gap of unknown length
 * 2513 2513: contig of 1014 bp in length
 * 2514 2514: gap of unknown length
 * 2515 2515: contig of 1014 bp in length
 * 2516 2516: gap of unknown length
 * 2517 2517: contig of 1014 bp in length
 * 2518 2518: gap of unknown length
 * 2519 2519: contig of 1014 bp in length
 * 2520 2520: gap of unknown length
 * 2521 2521: contig of 1014 bp in length
 * 2522 2522: gap of unknown length
 * 2523 2523: contig of 1014 bp in length
 * 2524 2524: gap of unknown length
 * 2525 2525: contig of 1014 bp in length
 * 2526 2526: gap of unknown length
 * 2527 2527: contig of 1014 bp in length
 * 2528 2528: gap of unknown length
 * 2529 2529: contig of 1014 bp in length
 * 2530 2530: gap of unknown length
 * 2531 2531: contig of 1014 bp in length
 * 2532 2532: gap of unknown length
 * 2533 2533: contig of 1014 bp in length
 * 2534 2534: gap of unknown length
 * 2535 2535: contig of 1014 bp in length
 * 2536 2536: gap of unknown length
 * 2537 2537: contig of 1014 bp in length
 * 2538 2538: gap of unknown length
 * 2539 2539: contig of 1014 bp in length
 * 2540 2540: gap of unknown length
 * 2541 2541: contig of 1014 bp in length
 * 2542 2542: gap of unknown length
 * 2543 2543: contig of 1014 bp in length
 * 2544 2544: gap of unknown length
 * 2545 2545: contig of 1014 bp in length
 * 2546 2546: gap of unknown length
 * 2547 2547: contig of 1014 bp in length
 * 2548 2548: gap of unknown length
 * 2549 2549: contig of 1014 bp in length
 * 2550 2550: gap of unknown length
 * 2551 2551: contig of 1014 bp in length
 * 2552 2552: gap of unknown length
 * 2553 2553: contig of 1014 bp in length
 * 2554 2554: gap of unknown length
 * 2555 2555: contig of 1014 bp in length
 * 2556 2556: gap of unknown length
 * 2557 2557: contig of 1014 bp in length
 * 2558 2558: gap of unknown length
 * 2559 2559: contig of 1014 bp in length
 * 2560 2560: gap of unknown length
 * 2561 2561: contig of 1014 bp in length
 * 2562 2562: gap of unknown length
 * 2563 2563: contig of 1014 bp in length
 * 2564 2564: gap of unknown length
 * 2565 2565: contig of 1014 bp in length
 * 2566 2566: gap of unknown length
 * 2567 2567: contig of 1014 bp in length
 * 2568 2568: gap of unknown length
 * 2569 2569: contig of 1014 bp in length
 * 2570 2570: gap of unknown length
 * 2571 2571: contig of 1014 bp in length
 * 2572 2572: gap of unknown length
 * 2573 2573: contig of 1014 bp in length
 * 2574 2574: gap of unknown length
 * 2575 2575: contig of 1014 bp in length
 * 2576 2576: gap of unknown length
 * 2577 2577: contig of 1014 bp in length
 * 2578 2578: gap of unknown length
 * 2579 2579: contig of 1014 bp in length
 * 2580 2580: gap of unknown length
 * 2581 2581: contig of 1014 bp in length
 * 2582 2582: gap of unknown length
 * 2583 2583: contig of 1014 bp in length
 * 2584 2584: gap of unknown length
 * 2585 2585: contig of 1014 bp in length
 * 2586 2586: gap of unknown length
 * 2587 2587: contig of 1014 bp in length
 * 2588 2588: gap of unknown length
 * 2589 2589: contig of 1014 bp in length
 * 2590 2590: gap of unknown length
 * 2591 2591: contig of 1014 bp in length
 * 2592 2592: gap of unknown length
 * 2593 2593: contig of 1014 bp in length
 * 2594 2594: gap of unknown length
 * 2595 2595: contig of 1014 bp in length
 * 2596 2596: gap of unknown length
 * 2597 2597: contig of 1014 bp in length
 * 2598 2598: gap of unknown length
 * 2599 2599: contig of 1014 bp in length
 * 2600 2600: gap of unknown length
 * 2601 2601: contig of 1014 bp in length
 * 2602 2602: gap of unknown length
 * 2603 2603: contig of 1014 bp in length
 * 2604 2604: gap of unknown length
 * 2605 2605: contig of 1014 bp in length
 * 2606 2606: gap of unknown length
 * 2607 2607: contig of 1014 bp in length
 * 2608 2608: gap of unknown length
 * 2609 2609: contig of 1014 bp in length
 * 2610 2610: gap of unknown length
 * 2611 2611: contig of 1014 bp in length
 * 2612 2612: gap of unknown length
 * 2613 2613: contig of 1014 bp in length
 * 2614 2614: gap of unknown length
 * 2615 2615: contig of 1014 bp in length
 * 2616 2616: gap of unknown length
 * 2617 2617: contig of 1014 bp in length
 * 2618 2618: gap of unknown length
 * 2619 2619: contig of 1014 bp in length
 * 2620 2620: gap of unknown length
 * 2621 2621: contig of 1014 bp in length
 * 2622 2622: gap of unknown length
 * 2623 2623: contig of 1014 bp in length
 * 2624 2624: gap of unknown length
 * 2625 2625: contig of 1014 bp in length
 * 2626 2626: gap of unknown length
 * 2627 2627: contig of 1014 bp in length
 * 2628 2628: gap of unknown length
 * 2629 2629: contig of 1014 bp in length
 * 2630 2630: gap of unknown length
 * 2631 2631: contig of 1014 bp in length
 * 2632 2632: gap of unknown length
 * 2633 2633: contig of 1014 bp in length
 * 2634 2634: gap of unknown length
 * 2635 2635: contig of 1014 bp in length
 * 2636 2636: gap of unknown length
 * 2637 2637: contig of 1014 bp in length
 * 2638 2638: gap of unknown length
 * 2639 2639: contig of 1014 bp in length
 * 2640 2640: gap of unknown length
 * 2641 2641: contig of 1014 bp in length
 * 2642 2642: gap of unknown length
 * 2643 2643: contig of 1014 bp in length
 * 2644 2644: gap of unknown length
 * 2645 2645: contig of 1014 bp in length
 * 2646 2646: gap of unknown length
 * 2647 2647: contig of 1014 bp in length
 * 2648 2648: gap of unknown length
 * 2649 2649: contig of 1014 bp in length
 * 2650 2650: gap of unknown length
 * 2651 2651: contig of 1014 bp in length
 * 2652 2652: gap of unknown length
 * 2653 2653: contig of 1014 bp in length
 * 2654 2654: gap of unknown length
 * 2655 2655: contig of 1014 bp in length
 * 2656 2656: gap of unknown length
 * 2657 2657: contig of 1014 bp in length
 * 2658 2658: gap of unknown length
 * 2659 2659: contig of 1014 bp in length
 * 2660 2660: gap of unknown length
 * 2661 2661: contig of 1014 bp in length
 * 2662 2662: gap of unknown length
 * 2663 2663: contig of 1014 bp in length
 * 2664 2664: gap of unknown length
 * 2665 2665: contig of 1014 bp in length
 * 2666 2666: gap of unknown length
 * 2667 2667: contig of 1014 bp in length
 * 2668 2668: gap of unknown length
 * 2669 2669: contig of 1014 bp in length
 * 2670 2670: gap of unknown length
 * 2671 2671: contig of 1014 bp in length
 * 2672 2672: gap of unknown length
 * 2673 2673: contig of 1014 bp in length
 * 2674 2674: gap of unknown length
 * 2675 2675: contig of 1014 bp in length
 * 2676 2676: gap of unknown length
 * 2677 2677: contig of 1014 bp in length
 * 2678 2678: gap of unknown length
 * 2679 2679: contig of 1014 bp in length
 * 2680 2680: gap of unknown length
 * 2681 2681: contig of 1014 bp in length
 * 2682 2682: gap of unknown length
 * 2683 2683: contig of 1014 bp in length
 * 2684 2684: gap of unknown length
 * 2685 2685: contig of 1014 bp in length
 * 2686 2686: gap of unknown length
 * 2687 2687: contig of 1014 bp in length
 * 2688 2688: gap of unknown length
 * 2689 2689: contig of 1014 bp in length
 * 2690 2690: gap of unknown length
 * 2691 2691: contig of 1014 bp in length
 * 2692 2692: gap of unknown length
 * 2693 2693: contig of 1


```

* 59081 59180: gap of unknown length
* 59181 60513: contig of 1333 bp in length
* 60514 60613: gap of unknown length
* 60614 62621: contig of 2008 bp in length
* 62622 62721: gap of unknown length
* 62722 64707: contig of 1986 bp in length
* 64708 64808: gap of unknown length
* 64809 66176: contig of 1369 bp in length
* 66177 66276: gap of unknown length
* 66277 68684: contig of 2408 bp in length
* 68685 70841: contig of 2057 bp in length
* 70842 70942: gap of unknown length
* 70942 72278: contig of 1337 bp in length
* 72279 73668: gap of unknown length
* 73669 73969: contig of 1490 bp in length
* 73970 75979: contig of 2011 bp in length
* 75980 76079: gap of unknown length
* 76080 78068: contig of 1989 bp in length
* 78069 78168: gap of unknown length
* 78169 79454: contig of 1286 bp in length
* 79455 79554: gap of unknown length
* 79555 80710: contig of 1155 bp in length
* 80710 82754: contig of 1945 bp in length
* 82755 82854: gap of unknown length
* 82855 85396: contig of 2542 bp in length
* 85397 85496: gap of unknown length
* 85497 88065: contig of 2569 bp in length
* 88066 88165: gap of unknown length
* 88166 90996: contig of 2831 bp in length
* 90997 91096: gap of unknown length
* 91097 92377: contig of 1481 bp in length
* 92378 94520: contig of 1843 bp in length
* 94521 94620: gap of unknown length

Query Match
Best Local Similarity 100.0%; Pred. No. 0;
Matches 634; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 827 GACCTCGCCTTCGCCACAACTACATTTGGGCACCTGACTCAGACGCTGGCGATAGCGGA 886
DB 57641 GACCTCGCCTTCGCCACAACTACATTTGGGCACCTGACTCAGACGCTGGCGATAGCGGA 57700

QY 887 CCACAGCTTCTACGGCCCGCGAGCCCGCTGTGCCCTGTGGGAGCTGGGAGCCCGGGAGG 946
DB 57701 CCACAGCTTCTACGGCCCGCGAGCCCGCTGTGCCCTGTGGGAGCTGGGAGCCCGGGAGG 57760

QY 947 GGGCTCCAGCGGACTGGGGCTCTATCTACTCCCGAGTTTCCCAAGCTGGTAGCCTGAG 1006
DB 57761 GGGCTCCAGCGGACTGGGGCTCTATCTACTCCCGAGTTTCCCAAGCTGGTAGCCTGAG 57820

QY 1007 CCCACAGCCTCATTTGAGGAGTTCCTGCCCTGCGCTGAGTCCCGAGCTCCCGATCCTGTCT 1066
DB 57821 CCCACAGCCTCATTTGAGGAGTTCCTGCCCTGCGCTGAGTCCCGAGCTCCCGATCCTGTCT 57880

QY 1067 GCTCCCGGCGACCTCGGTGTTCTCAGACTCTTGTGAAGGGCCCAACAGCGCCTGGGG 1126
DB 57881 GCTCCCGGCGACCTCGGTGTTCTCAGACTCTTGTGAAGGGCCCAACAGCGCCTGGGG 57940

QY 1127 GTGGCGCTGGCAGAAAGGAGGAGTACAGCTGCTCAATGGAAGTAGTGGAGGCA 1186
DB 57941 GTGGCGCTGGCAGAAAGGAGGAGTACAGCTGCTCAATGGAAGTAGTGGAGGCA 58000

QY 1187 CTCGAGCATCTCGCCCTCTCTGGCTTTTATTAGTCAGGTCCTGATTTAACAGGATTCG 1246
DB 58001 CTCGAGCATCTCGCCCTCTCTGGCTTTTATTAGTCAGGTCCTGATTTAACAGGATTCG 58060

QY 1247 CACAGTTCCTGCTGTGTGCGTGCACAAAGGACATTGACAGCTGATCTCTTTAAACC 1306
DB 58061 CACAGTTCCTGCTGTGTGCGTGCACAAAGGACATTGACAGCTGATCTCTTTAAACC 58120

```

```

QY 1307 TCCTCAGTGTGGCCACCTCAAACTCCCGCTCCAGCAGAGAGAGCGGTAGCCTAAATA 1366
DB 58121 TCCTCAGTGTGGCCACCTCAAACTCCCGCTCCAGCAGAGAGAGCGGTAGCCTAAATA 58180

QY 1367 GTTGGGAGACTCCCATATCTCTCTGGTGAATCCGCTCTTTTCAAACTCGCGGCTCCAA 1426
DB 58181 GTTGGGAGACTCCCATATCTCTCTGGTGAATCCGCTCTTTTCAAACTCGCGGCTCCAA 58240

QY 1427 CCACCGCTTTCTCCAGAGTGACCTTAATCCAGTGT 1460
DB 58241 CCACCGCTTTCTCCAGAGTGACCTTAATCCAGTGT 58274

RESULT 4
MMU76208 861 bp DNA linear ROD 05-FEB-1997
LOCUS Mus musculus neurogenin 3 (ngn3) gene, complete cds.
DEFINITION
ACCESSION U76208
VERSION U76208.1 GI:1815654
KEYWORDS
SOURCE Mus musculus.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 861)
AUTHORS Sommer,L., Ma,Q. and Anderson,D.J.
TITLE Sommer,L., Ma,Q. and Anderson,D.J.
JOURNAL neurogenins, a novel family of atonal-related bHLH transcription
MEDLINE factors, are putative mammalian neuronal determination genes that
PUBMED reveal progenitor cell heterogeneity in the developing CNS and PNS
97153585
9000438
2 (bases 1 to 861)
REFERENCE Sommer,L., Ma,Q. and Anderson,D.J.
AUTHORS Sommer,L., Ma,Q. and Anderson,D.J.
TITLE Direct Submission
JOURNAL Submitted (24-OCT-1996) Biology 216-76, California Institute of
Technology, Howard Hughes Medical Institute, Pasadena, CA 91125,
USA
COMMENT On Feb 5, 1997 this sequence version replaced gi:1666911.
FEATURES
source
1..861
/organism="Mus musculus"
/db_xref="taxon:10090"
gene 150..804
/gene="ngn3"
CDS 150..804
/function="bHLH transcription factor"
/note="NGN3"
/codon_start=1
/product="neurogenin 3"
/protein_id="AAC53029.1"
/db_xref="GI:1666912"
translation="MAPHPLDALTIQVSPETQOPFGASDHEVLSSNTPSPPTLIPR
DCSEAVGDCRGTSKRLARRGRNRKSELALSKORRRKKKANDRRNRHNLNSA
LDALRGVLPTEDDAKLKIETLREAHYIWAOTLRIADHSFYGPPEPVPCEGLS
PGGSGNGDWGSIYSPVSQAGNLSPTASLEEFFGLQVPSPSYLLPGALVFSDFL"
BASE COUNT 182 a 274 c 250 g 155 t
ORIGIN
Query Match
Best Local Similarity 100.0%; Pred. No. 1.5e-42; Length 861;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 762 GCGCTGGATCGCTCGCGGTGTCTCGCCACCTTCCCGATGACGCCAACTTACAAAG 821
DB 463 GCGCTGGATCGCTCGCGGTGTCTCGCCACCTTCCCGATGACGCCAACTTACAAAG 522

QY 822 ATCGAGACCCCTCGCTTCGCCCCACAACTACAT 853
DB 523 ATCGAGACCCCTCGCTTCGCCCCACAACTACAT 554

```

RESULT 5
 AF364300
 LOCUS
 DEFINITION
 AF364300
 ACCESSION
 AF364300.1 GI:13937128
 VERSION
 KEYWORDS
 SOURCE
 Mus musculus
 ORGANISM
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 REFERENCE
 1 (bases 1 to 1861)
 Lee, J., Smith, S., Watada, H., Lin, J., Scheel, D., Wang, J., Mirmira, R.
 and German, M.
 TITLE
 Regulation of the pancreatic pro-endocrine gene neurogenin3
 JOURNAL
 Diabetes (2001) In press
 REFERENCE
 2 (bases 1 to 1861)
 Schwitzgebel, V. and German, M.
 AUTHORS
 Direct Submission
 TITLE
 Submitted (26-MAR-2001) Hormone Research Institute, University of
 JOURNAL
 California San Francisco, 513 Parnassus Ave, HSW1090, San
 Francisco, CA 94145-0534, USA
 FEATURES
 Location/Qualifiers
 source
 1. .1861
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 mRNA
 <1093..>1737
 /product="neurogenin 3"
 CDS
 1093..1737
 /codon_start=1
 /product="neurogenin 3"
 /protein_id="AA050058.1"
 /db_xref="GI:13937129"
 translation="MAPHPLDALTIOVSPETQPPFGASDHVLSNSTPPSTLPIR
 DCSEAEVDCRGTSRKLRARRGRNRPKSELALSKQRSSRRKKANDRRNRHNLNSA
 LDALRGVLPFPDDAKLTETLRFAHNYIWTOTLRADHSFYGPPEVPVPGCELGS
 PGGSGNDWGSIIYSPYSQAGNLSPASLEEFPGLOVPSPSYLLPGALVFSDFL"
 BASE COUNT 397 a 560 c 537 g 367 t
 ORIGIN
 Query Match 6.3%; Score 92; DB 10; Length 1861;
 Best Local Similarity 100.0%; Pred. No. 1.5e-42;
 Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 762 GCGCTGGATGCGCTGCGGGTGCTCGCCACCTTCGCGGATGACGCCAACTTACAAAG 821
 Db 1396 GCGCTGGATGCGCTGCGGGTGCTCGCCACCTTCGCGGATGACGCCAACTTACAAAG 1455
 QY 822 ATCGAGACCTTCGCTTCGCCCACTACAT 853
 Db 1456 ATCGAGACCTTCGCTTCGCCCACTACAT 1487
 RESULT 6
 AF364300
 LOCUS
 DEFINITION
 AF364300
 ACCESSION
 AF364300.2 GI:11065737
 VERSION
 KEYWORDS
 SOURCE
 Mus musculus
 ORGANISM
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 REFERENCE
 1 (bases 1 to 5567)
 Cau, E., Gradwohl, G., Fode, C. and Guillemot, F.
 Mash1 activates a cascade of bHLH regulators in olfactory neuron
 progenitors
 JOURNAL
 Development 124 (8), 1611-1621 (1997)
 MEDLINE
 97261963
 PUBMED
 9108377
 REFERENCE
 2 (bases 1 to 5567)

AUTHORS
 Jacquemin, P., Durviaux, S.M., Jensen, J., Godfraind, C., Gradwohl, G.,
 Guillemot, F., Madsen, O.D., Carmeliet, P., Dewerchin, M., Collen, D.,
 Rousseau, G. and Lemaigre, F.P.
 TITLE
 Transcription factor hepatocyte nuclear factor 6 regulates
 pancreatic endocrine cell differentiation and controls expression
 of the proendocrine gene ngn3
 JOURNAL
 Mol. Cell. Biol. 20 (12), 4445-4454 (2000)
 MEDLINE
 20285449
 PUBMED
 10825208
 REFERENCE
 3 (bases 1 to 5567)
 Gradwohl, G.J.
 TITLE
 Direct Submission
 JOURNAL
 Submitted (04-NOV-1996) G.J. Gradwohl, IGBMC,
 CNRS-INSERM-Université Louis Pasteur, BP163, C.U. de Strasbourg,
 F-67404 ILLKIRCH cedex, FRANCE
 REMARK
 Revised by [4]
 REFERENCE
 4 (bases 1 to 5567)
 Lemaigre, F.P.
 TITLE
 Direct Submission
 JOURNAL
 Submitted (01-AUG-2000) Lemaigre F.P., Hormone and Metabolic
 Research Unit, Louvain University Medical School, Avenue Hippocrate
 75, box 7529, Brussels 1200, BELGIUM
 COMMENT
 On Oct 31, 2000 this sequence version replaced gi:1666087.
 FEATURES
 Location/Qualifiers
 source
 1. .5567
 /organism="Mus musculus"
 /strain="SV129"
 /db_xref="taxon:10090"
 /clone="lambda GEM12 clone 19C"
 /cell_line="D3"
 /cell_type="embryonic stem cells"
 /clone_lib="Genomic library No.62, MboI partial digests
 cloned in BamHI"
 /dev_stage="embryos"
 misc_binding 958
 /bound_moiety="HNF6"
 misc_binding 3692
 /bound_moiety="HNF6"
 gene 4145..5567
 /gene="MATH4B"
 4145..4151
 /gene="MATH4B"
 /note="putative"
 4923..5567
 /gene="MATH4B"
 /codon_start=1
 /product="neurogenin 3"
 /protein_id="CA070366.1"
 /db_xref="GI:1666088"
 /db_xref="SWISS-PROT:P70661"
 translation="MAPHPLDALTIOVSPETQPPFGASDHVLSNSTPPSTLPIR
 DCSEAEVDCRGTSRKLRARRGRNRPKSELALSKQRSSRRKKANDRRNRHNLNSA
 LDALRGVLPFPDDAKLTETLRFAHNYIWTOTLRADHSFYGPPEVPVPGCELGS
 PGGSGNDWGSIIYSPYSQAGNLSPASLEEFPGLOVPSPSYLLPGALVFSDFL"
 BASE COUNT 1271 a 1549 c 1564 g 1183 t
 ORIGIN
 Query Match 6.3%; Score 92; DB 10; Length 5567;
 Best Local Similarity 100.0%; Pred. No. 1.4e-42;
 Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 762 GCGCTGGATGCGCTGCGGGTGCTCGCCACCTTCGCGGATGACGCCAACTTACAAAG 821
 Db 5226 GCGCTGGATGCGCTGCGGGTGCTCGCCACCTTCGCGGATGACGCCAACTTACAAAG 5285
 QY 822 ATCGAGACCTTCGCTTCGCCCACTACAT 853
 Db 5286 ATCGAGACCTTCGCTTCGCCCACTACAT 5317
 RESULT 7
 AC109783/c
 LOCUS
 AC109783 138070 bp DNA linear HTG 07-FEB-2002

```

DEFINITION Mus musculus clone RP23-121F10, WORKING DRAFT SEQUENCE, 17
unordered pieces.
AC109783
AC109783.1 GI:18581594
VERSION HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS Mus musculus.
SOURCE Mus musculus.
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 138070)
McCombie,W.R., Baker,J.P., Ballija,V., Dedhia,N.N., de la
Bastide,M., Katzenberger,F., Kuit,K., King,L., Kirchoff,K.A.,
Miller,B., Muller,S., Nascimento,L.U., O'Shaughnessy,A.L.,
Preston,R.R., Santos,L., Spiegel,L.A., Palmer,L., Yang,C. and
Zutavern,T.
Mus Genomic Sequence
Unpublished
2 (bases 1 to 138070)
McCombie,W.R.
Direct Submission
Submitted (07-FEB-2002) Lita Annenberg Hazen Genome Sequencing
Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring
Harbor, NY 11724, USA
----- Genome Center -----
Center: Lita Annenberg Hazen Genome Center, Cold Spring Harbor
Laboratory
Laboratory
Center code: CSHL
Web site: http://www.cshl.org/geneseq
Contact: mcombie@cshl.org
----- Project Information
Project name: RP23-121F10
Clone name: RP23-121F10
Insert size: 173000; agarose-fp
Quality coverage: 4.00 in Q20 bases; agarose-fp
Quality coverage: 3.70 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 17 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 17785: contig of 17785 bp in length
* 17786 17874: gap of unknown length
* 17875 30356: contig of 12482 bp in length
* 30357 30444: gap of unknown length
* 30445 42306: contig of 11862 bp in length
* 42307 42394: gap of unknown length
* 42395 53598: contig of 11204 bp in length
* 53599 53686: gap of unknown length
* 53687 64355: contig of 10669 bp in length
* 64356 64443: gap of unknown length
* 64444 74017: contig of 9573 bp in length
* 74017 83366: gap of unknown length
* 83367 83454: gap of unknown length
* 83455 92355: contig of 8901 bp in length
* 92356 92444: gap of unknown length
* 92444 100821: contig of 8378 bp in length
* 100822 100909: gap of unknown length
* 100910 107529: contig of 6620 bp in length
* 107530 107617: gap of unknown length
* 107618 114066: contig of 6449 bp in length
* 114067 114154: gap of unknown length
* 114155 118873: contig of 4719 bp in length
* 118874 118961: gap of unknown length
* 118962 123619: contig of 4658 bp in length
* 123620 123707: gap of unknown length
* 123708 128240: contig of 4533 bp in length
* 128241 128328: gap of unknown length

```

```

DEFINITION Mus musculus chromosome UNK clone RP23-459M2, WORKING DRAFT
SEQUENCE, 36 unordered pieces.
AC127417 185806 bp DNA linear HTG 15-JUL-2002
Mus musculus
AC127417.1 GI:21759524
HTG; HTGS_PHASE1; HTGS_DRAFT.
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 185806)
McPherson,J.D. and Waterston,R.H.
The sequence of Mus musculus clone
Unpublished
2 (bases 1 to 185806)
McPherson,J.D. and Waterston,R.H.
Direct Submission
Submitted (15-JUL-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@wustl.wustl.edu
----- Project Information -----
Center project name: M.BA0459M02
----- Summary Statistics -----
Sequencing vector: M13; 0%
Sequencing vector: plasmid; 100%
Chemistry: Dye-primer ET; 0% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 173518 bases at least Q40
Consensus quality: 177276 bases at least Q30
Consensus quality: 179611 bases at least Q20
Insert size: 192000; agarose-fp
Insert size: 184535; sum-of-contigs
Quality coverage: 3.66 in Q20 bases; agarose-fp
Quality coverage: 3.85 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 36 contigs. The true order of the pieces
* is not known and their order in this sequence record is

```

```

FEATURES
source
1..138070
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="RP23-121F10"
BASE COUNT 35731 a 33657 c 33303 g 33954 t 1425 others
ORIGIN
Query Match 6.3% Score 92; DB 2; Length 138070;
Best Local Similarity 100.0%; Pred. No. 1.2e-42;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 762 GCCTGGATCGCTGCGGTGCTCTGCCACCTTCCCGATGACGCCAACTTACAAG 821
Db 110917 GCCTGGATCGCTGCGGTGCTCTGCCACCTTCCCGATGACGCCAACTTACAAG 110858
QY 822 ATGAGACCTTGGCTTCCGCCACACTACAT 853
Db 110857 ATGAGACCTTGGCTTCCGCCACACTACAT 110826
RESULT 8
AC127417
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
AUTHORS
TITLE
JOURNAL
COMMENT

```

- * arbitrary. Gaps between the contigs are represented as
- * runs of N, but the exact sizes of the gaps are unknown.
- * This record will be updated with the finished sequence
- * as soon as it is available and the accession number will
- * be preserved.

*	1	1042:	contig	of 1042	bp in length
*	1043	1142:	gap of	unknown	length
*	1143	2216:	contig of	1074	bp in length
*	2217	2316:	gap of	unknown	length
*	2317	3838:	contig of	1522	bp in length
*	3839	3938:	gap of	unknown	length
*	3939	5356:	contig of	1458	bp in length
*	5397	5496:	gap of	unknown	length
*	5497	6874:	contig of	1378	bp in length
*	6875	6974:	gap of	unknown	length
*	6975	8481:	contig of	1507	bp in length
*	8482	8581:	gap of	unknown	length
*	8582	9874:	contig of	1293	bp in length
*	9875	9974:	gap of	unknown	length
*	9975	11483:	contig of	1509	bp in length
*	11484	11583:	gap of	unknown	length
*	11584	13834:	contig of	2251	bp in length
*	13835	13934:	gap of	unknown	length
*	13935	16035:	contig of	2101	bp in length
*	16036	16135:	gap of	unknown	length
*	16136	19100:	contig of	2965	bp in length
*	19101	19200:	gap of	unknown	length
*	19201	22184:	contig of	2984	bp in length
*	22185	22884:	gap of	unknown	length
*	22885	25287:	contig of	3003	bp in length
*	25288	25387:	gap of	unknown	length
*	25388	28431:	contig of	3044	bp in length
*	28432	28531:	gap of	unknown	length
*	28532	31409:	contig of	2878	bp in length
*	31410	31509:	gap of	unknown	length
*	31510	35320:	contig of	3811	bp in length
*	35321	35420:	gap of	unknown	length
*	35421	38808:	contig of	3388	bp in length
*	38809	38908:	gap of	unknown	length
*	38909	43846:	contig of	4938	bp in length
*	43847	43946:	gap of	unknown	length
*	43947	49171:	contig of	5225	bp in length
*	49172	49271:	gap of	unknown	length
*	49272	53157:	contig of	3886	bp in length
*	53158	53257:	gap of	unknown	length
*	53258	57731:	contig of	4474	bp in length
*	57732	57831:	gap of	unknown	length
*	57932	62660:	contig of	4829	bp in length
*	62661	62760:	gap of	unknown	length
*	62761	68243:	contig of	5483	bp in length
*	68244	68343:	gap of	unknown	length
*	68344	73047:	contig of	4704	bp in length
*	73048	73147:	gap of	unknown	length
*	73148	78565:	contig of	5418	bp in length
*	78566	78665:	gap of	unknown	length
*	78666	85726:	contig of	7061	bp in length
*	85727	85826:	gap of	unknown	length
*	85827	91985:	contig of	6159	bp in length
*	91986	92085:	gap of	unknown	length
*	92086	97984:	contig of	5899	bp in length
*	97985	98084:	gap of	unknown	length
*	98085	105338:	contig of	7254	bp in length
*	105339	105438:	gap of	unknown	length
*	105439	113368:	contig of	7930	bp in length
*	113369	113468:	gap of	unknown	length
*	113469	120753:	contig of	7285	bp in length
*	120754	120853:	gap of	unknown	length
*	120854	129342:	contig of	8489	bp in length
*	129343	129442:	gap of	unknown	length
*	129443	138908:	contig of	9556	bp in length
*	138999	139098:	gap of	unknown	length
*	139099	149579:	contig of	10481	bp in length
*	149580	149679:	gap of	unknown	length

	* 149680	162343:	contig of 12664 bp in length.
	* 162344	162443:	gap of unknown length
	* 162444	185806:	contig of 23363 bp in length.
FEATURES	Location/Qualifiers		
source	1..185806		
	/organism="Mus musculus"		
	/db_xref="taxon:10090"		
	/chromosome="UNK"		
	/clone="RP23-459M2"		
	1..1042		
misc_feature	/note="assembly_name:Contig24"		
	1143..2216		
misc_feature	/note="assembly_name:Contig25"		
	2317..3838		
misc_feature	/note="assembly_name:Contig26"		
	3939..5396		
misc_feature	/note="assembly_name:Contig27"		
	5497..6874		
misc_feature	/note="assembly_name:Contig28"		
	6975..8481		
misc_feature	/note="assembly_name:Contig29"		
	8582..9874		
misc_feature	/note="assembly_name:Contig30"		
	9975..11483		
misc_feature	/note="assembly_name:Contig31"		
	11584..13834		
misc_feature	/note="assembly_name:Contig32"		
	13935..16035		
misc_feature	/note="assembly_name:Contig33"		
	16136..19100		
misc_feature	/note="assembly_name:Contig34"		
	19201..22184		
misc_feature	/note="assembly_name:Contig35"		
	22285..25287		
misc_feature	/note="assembly_name:Contig36"		
	25388..28431		
misc_feature	/note="assembly_name:Contig37"		
	28532..33409		
misc_feature	/note="assembly_name:Contig38"		
	31510..35320		
misc_feature	/note="assembly_name:Contig39"		
	35421..38808		
misc_feature	/note="assembly_name:Contig40"		
	38909..43846		
misc_feature	/note="assembly_name:Contig41"		
	43947..49171		
misc_feature	/note="assembly_name:Contig42"		
	49272..53157		
misc_feature	/note="assembly_name:Contig43"		
	53258..57731		
misc_feature	/note="assembly_name:Contig44"		
	57832..62660		
misc_feature	/note="assembly_name:Contig45"		
	62761..68243		
misc_feature	/note="assembly_name:Contig46"		
	68344..73047		
misc_feature	/note="assembly_name:Contig47"		
	73148..78565		
misc_feature	/note="assembly_name:Contig48"		
	78666..85726		
misc_feature	/note="assembly_name:Contig49"		
	85827..91985		
misc_feature	/note="assembly_name:Contig50"		
	92086..97984		
misc_feature	/note="assembly_name:Contig51		
	clone_end.SP6		
	vector.side:left		
misc_feature	98085..105338		
	/note="assembly_name:Contig52"		
misc_feature	105439..113368		
	/note="assembly_name:Contig53"		
misc_feature	113469..120753		
	/note="assembly_name:Contig54"		

misc_feature 120854..129342
 /note="assembly_name:Contig55"
 misc_feature 129443..138998
 /note="assembly_name:Contig56"
 misc_feature 139099..149579
 /note="assembly_name:Contig57"
 misc_feature 149680..162343
 /note="assembly_name:Contig58"
 misc_feature 162444..185806

Query Match 6.3%; Score 92; DB 2; Length 185806;
 Best Local Similarity 100.0%; Pred. No. 1.2e-42;
 Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 762 GCCTGGATGCGTGGCGGTGTCCTGCCACCTTCCCGATGACGCCCAACTTCAAAAG 821
 |||||
 DB 159127 GCCTGGATGCGTGGCGGTGTCCTGCCACCTTCCCGATGACGCCCAACTTCAAAAG 159186
 |||||
 QY 822 ATGAGACCTCGCTTGGCCCAACTACAT 853
 |||||
 DB 159187 ATGAGACCTCGCTTGGCCCAACTACAT 159218
 |||||

RESULT 9
 AF303002
 LOCUS AF303002 6123 bp DNA linear PRI 13-NOV-2001
 DEFINITION Homo sapiens neurogenin 2 gene, partial cds.
 ACCESSION AF303002
 VERSION AF303002.1 GI:11875763
 KEYWORDS
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 6123)
 AUTHORS Simmons,A.D., Horton,S., Abney,A.L. and Johnson,J.E.
 TITLE Neurogenin2 expression in ventral and dorsal spinal neural tube
 JOURNAL progenitor cells is regulated by distinct enhancers
 MEDLINE Dev Biol. 229 (2), 327-339 (2001)
 PUBMED 21077024
 REFERENCE 2 (bases 1 to 6123)
 AUTHORS Simmons,A.D., Horton,S., Abney,A.L. and Johnson,J.E.
 TITLE Direct Submission
 JOURNAL Submitted (06-SEP-2000) Center for Basic Neuroscience - N44.146, UT
 Southwestern Medical Center, 5323 Harry Hines Boulevard, Dallas, TX
 75390-9111, USA

FEATURES
 source
 1..6123
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 mRNA
 <1..>754
 /product="neurogenin 2"
 CDS
 <1..754
 /note="NGN2"
 /codon_start=2
 /product="neurogenin 2"
 /protein_id="AA040770.1"
 /db_xref="GI:11875764"
 /translation="ASPALAALTPLSSADEEEEEEPGASGGARRORGAEGQGARGG
 VAAGGCGPRLGLVHDCRPRARAVSRAGTAETVQRIKTRRLKANNRERNR
 MINLNALDALREVLPFPEDAKLTKIETLRAHNYIWALTETLRLADHCGGGGLP
 GALTSEAVLUSFGGASASGDSFSPASTWCSINSPAPSSVSNSTSPISCTLSLSP
 ASPAGSDMDYQWPPPPDKRYAPHPDIARDCI"
 BASE COUNT 1484 a 1536 c 1507 g 1596 t
 ORIGIN
 1484 a 1536 c 1507 g 1596 t

Query Match 2.4%; Score 35; DB 9; Length 6123;
 Best Local Similarity 100.0%; Pred. No. 1.7e-08;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 819 AAGATCGAGACCTGGCTTCGCCCAACTACAT 853
 |||||

DB 383 AAGATCGAGACCTGGCTTCGCCCAACTACAT 417

RESULT 10
 AC079846_3/c
 WPCOMMENT
 Sequence split into 4 fragments LOCUS AC079846 Accession AC079846
 Fragment Name Begin End
 AC079846_0 1 110000
 AC079846_1 100001 210000
 AC079846_2 200001 310000
 AC079846_3 300001 391531
 Continuation (4 of 4) of AC079846 from base 300001 (AC079846 Homo sapiens chromosome

Query Match 2.4%; Score 35; DB 2; Length 91531;
 Best Local Similarity 100.0%; Pred. No. 1.4e-08;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 819 AAGATCGAGACCTGGCTTCGCCCAACTACAT 853
 |||||
 DB 86995 AAGATCGAGACCTGGCTTCGCCCAACTACAT 86961
 |||||

RESULT 11
 AC023886/c
 LOCUS AC023886 179697 bp DNA linear PRI 20-MAR-2002
 DEFINITION Homo sapiens BAC clone RP11-402J6 from 4, complete sequence.
 ACCESSION AC023886
 VERSION AC023886.7 GI:19482381
 KEYWORDS HTG.
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 179697)
 AUTHORS Sulston,J.E. and Waterston,R.
 TITLE Toward a complete human genome sequence
 JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
 MEDLINE 99063792
 PUBMED 9847074

REFERENCE 2 (bases 1 to 179697)
 AUTHORS Radionenko,M. and Abbott,A.
 TITLE The sequence of Homo sapiens BAC clone RP11-402J6
 JOURNAL Unpublished (2001)

REFERENCE 3 (bases 1 to 179697)
 AUTHORS Waterston,R.H.
 TITLE Direct Submission
 JOURNAL Submitted (18-FEB-2000) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA

REFERENCE 4 (bases 1 to 179697)
 AUTHORS Waterston,R.H.
 TITLE Direct Submission

JOURNAL Submitted (15-MAR-2002) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA

REFERENCE 5 (bases 1 to 179697)
 AUTHORS Waterston,R.
 TITLE Direct Submission

JOURNAL Submitted (20-MAR-2002) Department of Genetics, Washington
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 On Mar 15, 2002 this sequence version replaced gi:17352441.

COMMENT
 ----- Genome Center
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: http://genome.wustl.edu/gsc
 Contact: sapiens@wustl.edu
 ----- Summary Statistics

 Center project name: H_NH0402J06

NOTICE: This sequence may not represent the entire insert of this
 clone. It may be shorter because we only sequence overlapping

clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osogawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tatenoe, M., Cataneese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org> and coworkers at <http://www.chori.org> VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is AC004049, 2000 bp overlap; the clone sequenced to the right is RP11-148B6, Actual end of this clone is at base position 179697 of RP11-402J6.

FEATURES

source	
1. .179697	Location/Qualifiers
/organism="Homo sapiens"	
/db_xref="taxon:9606"	
/chromosome="4"	
/map="4"	
/clone="RP11-402J6"	
/clone_lib="RPCI-11"	
1. .749	
/rpt_family="L1"	
734. .3067	
/rpt_family="L1"	
3104. .3444	
/rpt_family="MaLR"	
3447. .4289	
/rpt_family="L1"	
4290. .4369	
/rpt_family="(TA)n"	
4370. .4797	
/rpt_family="L1"	
5022. .5194	
/rpt_family="L2"	
5214. .5628	
/rpt_family="L1"	
5710. .5729	
/rpt_family="(TA)n"	
5749. .5863	
/rpt_family="(TA)n"	
5883. .6263	
/rpt_family="ERV1"	
9901. .10151	
/rpt_family="ERV1"	
10152. .10310	
/rpt_family="L1"	
10311. .10400	
/rpt_family="Mariner"	
10401. .10816	
/rpt_family="L1"	
10967. .12362	
/rpt_family="L1"	
12366. .12653	
/rpt_family="Alu"	
12695. .13940	
/rpt_family="L1"	
13965. .14272	
/rpt_family="Alu"	
14274. .14795	
/rpt_family="ERV1"	
14806. .15392	
/rpt_family="L1"	
21504. .22112	
/rpt_family="L1"	
22131. .23166	
/rpt_family="L1"	
23246. .23386	
/rpt_family="MaLR"	
23511. .23654	
/rpt_family="MIR"	
23655. .25486	
/rpt_family="L1"	
25487. .25636	
/rpt_family="MIR"	
25814. .26124	
/rpt_family="Alu"	
26444. .26592	
/rpt_family="MIR"	
27423. .27713	
/rpt_family="Alu"	
27716. .27739	
/rpt_family="AT-rich"	
27773. .27795	
/rpt_family="AT-rich"	
28020. .28429	
/rpt_family="ERV1"	
29646. .29668	
/rpt_family="AT-rich"	
29727. .29871	
/rpt_family="L1"	
29875. .30114	
/rpt_family="Alu"	
30134. .30267	
/rpt_family="L1"	
30268. .30549	
/rpt_family="L1"	
30550. .30883	
/rpt_family="MaLR"	
32461. .32791	
/rpt_family="MaLR"	
32792. .33487	
/rpt_family="L1"	
33488. .33847	
/rpt_family="MaLR"	
33848. .34354	
/rpt_family="L1"	
34358. .34379	
/rpt_family="(CATATA)n"	
34381. .34502	
/rpt_family="Alu"	
34503. .34556	
/rpt_family="L1"	
34557. .34859	
/rpt_family="Alu"	
34860. .35041	
/rpt_family="L1"	
35219. .35676	
/rpt_family="L2"	
35974. .36042	
/rpt_family="L2"	
36833. .37144	
/rpt_family="Alu"	
36833. .36848	
/note="match to EST AI537116 (NID:94451251) to15h09.xl"	
37147. .37410	
/rpt_family="Alu"	
38154. .38424	

```

QY 672 AAGCGAGTTGGCACTGAGCAAGCAGCGAG 703
Db 535 AAGCGAGTTGGCACTGAGCAAGCAGCGAG 566

RESULT 13
AF234829 5340 bp DNA linear PRI 19-OCT-2001
DEFINITION Homo sapiens neurogenin 3 gene, complete cds.
ACCESSION AF234829
VERSION AF234829.1 GI:13183002
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 5340)
AUTHORS del Bosque-Plata, L., Lin, J., Horikawa, Y., Schwarz, P. E., Cox, N. J.,
Iwasaki, N., Ogata, M., Iwamoto, Y., German, M. S. and Bell, G. I.
TITLE Mutations in the coding region of the neurogenin 3 gene (NEUROG3)
are not a common cause of maturity-onset diabetes of the young in
Japanese subjects
JOURNAL Diabetes 50 (3), 694-696 (2001)
MEDLINE 21140923
PUBMED 11248894
REFERENCE 2 (bases 1 to 5340)
AUTHORS Lin, J. and German, M.
TITLE Direct Submission
JOURNAL Submitted (15-FEB-2000) Hormone Research Institute, University of
California San Francisco, 513 Parnassus Ave., San Francisco, CA
94143-0534, USA
FEATURES
source Location/Qualifiers
1..5340
/organism="Homo sapiens"
/db_xref="taxon:9606"
<3022..>3666
/product="neurogenin 3"
3022..3666
/codon_start=1
/product="neurogenin 3"
/protein_id="AAK15022.1"
/db_xref="GI:13183003"
/translati="MTPOPSGAPTQVTRERSFRASEDEVTCPTSPAPSPTRTGRG
NCAEAGGCGRGAPEKLARRGGRSRPKSALSKORRRKANDRNRNMHNLNLA
LDALRGVLPFPDALKTKTETFEAHNYIWLQTLRIADHSILYALEPAPHGCELG
SPGSGDWSGLSPVSGASLSPASLEERPGLLGATYSACLSPGSLAFSDFL"
BASE COUNT 1215 a 1500 c 1514 g 1111 t
ORIGIN
Query Match 2.2%; Score 32; DB 9; Length 5340;
Best Local Similarity 100.0%; Pred. No. 1e-06;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 672 AAGCGAGTTGGCACTGAGCAAGCAGCGAG 703
Db 3235 AAGCGAGTTGGCACTGAGCAAGCAGCGAG 3266

RESULT 14
AL450311/c 165110 bp DNA linear PRI 12-JUL-2001
LOCUS Human DNA sequence from clone RP11-343J3 on chromosome 10, complete
DEFINITION sequence.
ACCESSION AL450311
VERSION AL450311.11 GI:14626972
KEYWORDS human.
SOURCE Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 165110)
AUTHORS Howden, P.

```


TITLE

Submitted (12-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk

COMMENT

On Jul 8, 2001 this sequence version replaced g1:14575291.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em: EMBL; SW:
SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP
database can be found at

<http://www.sanger.ac.uk/Projects/C-elegans/wormpep> This sequence
was generated from part of bacterial clone contigs of human
chromosome 10, constructed by the Sanger Centre Chromosome 10
Mapping Group. Further information can be found at

<http://www.sanger.ac.uk/HGP/Chr10>
RP11-34333 is from the library RP11-11.2 constructed by the group
of Pieter de Jong. For further details see
<http://www.chori.org/bacpac/home.htm>
VECTOR: pBACe3.6

This sequence is the entire insert of clone RP11-34333 The true
left end of clone RP11-242G20 is at 139955 in this sequence. The
true right end of clone RP11-404C6 is at 588 in this sequence.

FEATURES

source

1. 165110
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-34333"
/clone_lib="RP11-11.2"
7. 147

repeat_region

/note="THE1C repeat: matches 2. 142 of consensus"

repeat_region

/note="MIR repeat: matches 2. 148 of consensus"

repeat_region

/note="26 copies 2 mer gt 98% conserved"

repeat_region

/note="AluSg repeat: matches 1. 310 of consensus"

repeat_region

/note="AluSx repeat: matches 1. 134 of consensus"

repeat_region

/note="AluSx repeat: matches 118. 293 of consensus"

repeat_region

/note="L1MEC repeat: matches 272. 1095 of consensus"

repeat_region

/note="L1MEC repeat: matches 1168. 2367 of consensus"

repeat_region

/note="MIR repeat: matches 25. 184 of consensus"

repeat_region

/note="MIR repeat: matches 65. 194 of consensus"

repeat_region

/note="AluSx repeat: matches 1. 294 of consensus"

repeat_region

/note="20 copies 2 mer tc 95% conserved"

repeat_region

/note="65 copies 2 mer at 90% conserved"

repeat_region

/note="M1T1C repeat: matches 1. 466 of consensus"

repeat_region

/note="AluSg repeat: matches 1. 301 of consensus"

/note="AluSx repeat: matches 243. 299 of consensus"

repeat_region

14843. 14943
/note="MER86 repeat: matches 9. 109 of consensus"

repeat_region

15404. 15613
/note="MIR repeat: matches 3. 213 of consensus"

repeat_region

16226. 16705
/note="MIR repeat: matches 60. 140 of consensus"

repeat_region

16928. 16993
/note="33 copies 2 mer gg 66% conserved"

repeat_region

17476. 17569
/note="MER81 repeat: matches 2. 114 of consensus"

repeat_region

17719. 18069
/note="L1MC4 repeat: matches 7617. 7977 of consensus"

repeat_region

18117. 18345
/note="MIR repeat: matches 7. 262 of consensus"

repeat_region

19169. 19400
/note="MIR repeat: matches 5. 239 of consensus"

repeat_region

19436. 19627
/note="L1MC5 repeat: matches 7728. 7917 of consensus"

repeat_region

19642. 19716
/note="MIR repeat: matches 48. 131 of consensus"

repeat_region

19969. 20010
/note="21 copies 2 mer tg 100% conserved"

repeat_region

20623. 20704
/note="41 copies 2 mer gt 85% conserved"

repeat_region

20738. 20773
/note="U2 repeat: matches 1. 36 of consensus"

repeat_region

20802. 20863
/note="M1T1J repeat: matches 1. 62 of consensus"

repeat_region

20889. 21263
/note="M1T1F repeat: matches 188. 541 of consensus"

repeat_region

21463. 21618
/note="MIR repeat: matches 46. 192 of consensus"

repeat_region

22019. 22326
/note="AluJb repeat: matches 1. 306 of consensus"

repeat_region

22381. 22564
/note="L2 repeat: matches 2453. 2629 of consensus"

repeat_region

22896. 23174
/note="M1T1J repeat: matches 117. 413 of consensus"

repeat_region

23215. 23346
/note="MIR repeat: matches 48. 188 of consensus"

repeat_region

23388. 23532
/note="L2 repeat: matches 2097. 2230 of consensus"

repeat_region

23533. 23837
/note="AluJb repeat: matches 1. 303 of consensus"

repeat_region

23838. 24137
/note="L2 repeat: matches 1754. 2097 of consensus"

repeat_region

24291. 24581
/note="AluSx repeat: matches 1. 300 of consensus"

repeat_region

24653. 24850
/note="MIR repeat: matches 1. 200 of consensus"

repeat_region

26189. 26685
/note="CpG island"

repeat_region

29453. 30918
/note="CpG island"

repeat_region

31084. 31156
/note="MIR repeat: matches 65. 138 of consensus"

repeat_region

31618. 31876
/note="AluJb repeat: matches 29. 275 of consensus"

repeat_region

32767. 32830
/note="MIR repeat: matches 76. 139 of consensus"

repeat_region

33050. 33178
/note="43 copies 3 mer tcc 72% conserved"

repeat_region

35112. 36201
/note="CpG island"

repeat_region

36551. 36604
/note="27 copies 2 mer ac 94% conserved"

repeat_region

37585. 38254
/note="L1P15 repeat: matches 5480. 6157 of consensus"

repeat_region

38265. 38561
/note="AluSx repeat: matches 1. 300 of consensus"

```

repeat_region 39214..39423
/note="MIR repeat: matches 3. .219 of consensus"
repeat_region 39489..39801
/note="AluX repeat: matches 1. .292 of consensus"
repeat_region 40050..40189
/note="MIR repeat: matches 1. .144 of consensus"
repeat_region 41057..41290
/note="MIR repeat: matches 8. .240 of consensus"
repeat_region 41296..41460
/note="L2 repeat: matches 2569. .2730 of consensus"
repeat_region 41506..41944
/note="L2 repeat: matches 1916. .2416 of consensus"
repeat_region 42388..42698
/note="L2 repeat: matches 1448. .1779 of consensus"
repeat_region 44193..44579
/note="THE1C repeat: matches 1. .371 of consensus"
repeat_region 44600..44956
/note="LNR16A repeat: matches 90. .445 of consensus"
repeat_region 45240..45300
/note="MER58A repeat: matches 37. .97 of consensus"
repeat_region 45798..45909
/note="LNR41 repeat: matches 90. .192 of consensus"
repeat_region 46826..46871
/note="23 copies 2 mer gt 97% conserved"
repeat_region 50684..50969
/note="AluSq repeat: matches 1. .287 of consensus"
repeat_region 50980..51291
/note="AluX repeat: matches 1. .312 of consensus"
repeat_region 52222..52519
/note="AluX repeat: matches 1. .300 of consensus"
repeat_region 54065..54260
/note="LIM4 repeat: matches 3865. .4055 of consensus"
repeat_region 54261..54432
/note="FAM repeat: matches 2. .167 of consensus"
repeat_region 54433..54629
/note="LIM4 repeat: matches 3652. .3865 of consensus"
repeat_region 54648..54862
/note="LNR41 repeat: matches 11. .217 of consensus"
repeat_region 54863..55236
/note="MT1A1 repeat: matches 1. .365 of consensus"
repeat_region 55237..55700
/note="LNR41 repeat: matches 217. .716 of consensus"

Query Match 2.28; Score 32; DB 9; Length 165110;
Best Local Similarity 100.0%; Pred. No. 8.6e-07;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 672 AAGAGCGAGTTGGCAGTGCAGCAGCGAGCG 703
|||||
Db 30112 AAGAGCGAGTTGGCAGTGCAGCAGCGAGCG 30081

RESULT 15
AC021954/c
LOCUS
DEFINITION
Homo sapiens chromosome 10 clone RP11-57E12 map 10, WORKING DRAFT
SEQUENCE, 24 unordered pieces.
ACCESSION
AC021954
VERSION
AC021954.3 GI:7417809
KEYWORDS
HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 173341)
AUTHORS
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE
Homo sapiens chromosome 10, clone RP11-57E12
JOURNAL
Unpublished
REFERENCE
2 (bases 1 to 173341)
AUTHORS
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Bada,F.,
Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
Choepeil,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,

```

```

DeArellano,K., Dewar,K., Domino,M., Domino,M., Doyle,M., Fenestor,J.,
Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Gage,D., Galagan,J.,
Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howard,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lander,T., Lehoczy,J., Levine,R., Liu,C., Liu,C., Liu,C., Locke,K.,
Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., Olivari,T.M., Peterson,K.,
Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
Roy,A., Santos,R., Severi,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
Zimmer,A. and Zody,M.
Direct Submission
Submitted (22-JAN-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 173341)
REFERENCE
AUTHORS
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Bada,F.,
Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campoliano,A., Castle,A., Choepeil,Y., Colangelo,M., Collins,S.,
Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howard,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Laroque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Liu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neil,D., Olivari,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severi,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Willson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 5, 2000 this sequence version replaced gi:6984451.
All repeats were identified using RepeatMasker:
Snit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www.seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L5931
Center clone name: 57_E12
----- Summary Statistics
Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 161190 bases at least Q40
Consensus quality: 166837 bases at least Q30
Consensus quality: 168995 bases at least Q20
Insert size: 176000; agarose-fp
Insert size: 171041; sum-of-contigs
Quality coverage: 3.7 in Q20 bases; agarose-fp
Quality coverage: 3.8 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 24 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

```

* 1 1574: contig of 1574 bp in length
* 1575 1674: gap of 100 bp
* 1675 3043: contig of 1369 bp in length
* 3044 3143: gap of 100 bp
* 3144 5889: contig of 2746 bp in length
* 5890 5989: gap of 100 bp
* 5990 8979: contig of 2990 bp in length
* 8980 9079: gap of 100 bp
* 9080 13674: contig of 4595 bp in length
* 13675 13774: gap of 100 bp
* 13775 18931: contig of 5057 bp in length
* 18932 18933: gap of 100 bp
* 18933 23526: contig of 4595 bp in length
* 23527 23626: gap of 100 bp
* 23627 27386: contig of 3760 bp in length
* 27387 27486: gap of 100 bp
* 27487 32572: contig of 5086 bp in length
* 32573 32672: gap of 100 bp
* 32673 38632: contig of 5960 bp in length
* 38633 38732: gap of 100 bp
* 38733 43735: contig of 5003 bp in length
* 43736 43835: gap of 100 bp
* 43836 49020: contig of 5185 bp in length
* 49021 49120: gap of 100 bp
* 49121 53660: contig of 4540 bp in length
* 53661 53760: gap of 100 bp
* 53761 59544: contig of 5784 bp in length
* 59545 59644: gap of 100 bp
* 59645 66395: contig of 6751 bp in length
* 66396 66495: gap of 100 bp
* 66496 74696: contig of 8201 bp in length
* 74697 74796: gap of 100 bp
* 74797 81326: contig of 6530 bp in length
* 81327 81426: gap of 100 bp
* 81427 89446: contig of 8020 bp in length
* 89447 89546: gap of 100 bp
* 89547 98459: contig of 6913 bp in length
* 98460 98559: gap of 100 bp
* 98560 106498: contig of 9939 bp in length
* 106499 106598: gap of 100 bp
* 106599 116987: contig of 10388 bp in length
* 116988 117086: gap of 100 bp
* 117087 128890: contig of 11804 bp in length
* 128891 128990: gap of 100 bp
* 128991 147290: contig of 18300 bp in length
* 147291 147390: gap of 100 bp
* 147391 173341: contig of 25951 bp in length.

```

FEATURES

Source

```

1..173341
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/map="10"
/clone="RP11-57E12"
/cloneLib="RPC1-11 Human Male BAC"
1..1574
/note="assembly_fragment"
1675..3043
/note="assembly_fragment"
3144..5889
/note="assembly_fragment"
5990..8979
/note="assembly_fragment"
9080..13674
/note="assembly_fragment"
13775..18931
/note="assembly_fragment"
18932..23526
/note="assembly_fragment"
23627..27386
/note="assembly_fragment"
clone_end:SP6
vector_side:left"

```

```

misc_feature 27487..32572
/note="assembly_fragment"
misc_feature 32673..38632
/note="assembly_fragment"
misc_feature 38733..43735
/note="assembly_fragment"
misc_feature 43836..49020
/note="assembly_fragment"
misc_feature 49121..53660
/note="assembly_fragment"
misc_feature 53761..59544
/note="assembly_fragment"
misc_feature 59645..66395
/note="assembly_fragment"
misc_feature 66496..74696
/note="assembly_fragment"
misc_feature 74797..81326
/note="assembly_fragment"
misc_feature 81427..89446
/note="assembly_fragment"
misc_feature 89547..96459
/note="assembly_fragment"
misc_feature 96560..106498
/note="assembly_fragment"
misc_feature 106599..116986
/note="assembly_fragment"
misc_feature 117087..128890
/note="assembly_fragment"
clone_end:T7
vector_side:right"
misc_feature 128991..147290
/note="assembly_fragment"
misc_feature 147391..173341
/note="assembly_fragment"
BASE COUNT 44173 a 40928 c 40432 g 45501 t 2307 others
Query Match 2.2% Score 32; DB 2; Length 173341;
Best Local Similarity 100.0%; Pred. No. 8.6e-07;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 672 AAGACCGAGTTGGCACTGAGCAAGCAGCGACG 703
|||||
Db 6552 AAGACCGAGTTGGCACTGAGCAAGCAGCGACG 6521
Search completed: April 8, 2003, 22:56:52
Job time : 5752 secs

```

GenCore version 5.1.4_p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 8, 2003, 18:58:46 ; Search time 374 Seconds

(without alignments)
8791.223 Million cell updates/sec

Title: US-09-595-947C-1

Perfect score: 1460

Sequence: 1 gcaggtagcagagagcag.....agagtagcactaatccagtg 1460

Scoring table:

OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 2185239 seqs, 1125999159 residues

Word size : 15

Total number of hits satisfying chosen parameters: 4152

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_101002.*

```

1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1460	100.0	1491	19 AAV42512	cdNA encoding a no
2	92	6.3	804	19 AAV27050	Mouse neurogenin 3
3	92	6.3	804	21 AAZ51981	Murine neurogenin-3
4	92	6.3	861	22 AAF27266	Mouse neurogenin
5	92	6.3	1861	21 AAC61090	Murine neurogenin
6	92	6.3	5567	22 AAF27254	Mouse atonal homol
7	65	4.5	65	24 ABN31392	Rat spliced trans
8	35	2.4	65	24 ABN57521	Mouse spliced tran
9	32	2.2	5340	21 AAC61089	Human neurogenin 3

	10	29	2.0	428	22	AA333797	Human cdNA encodin
c	11	26	1.8	26	24	ABT03700	Human Neurogenin-3
	12	25	1.7	25	19	AAV42515	PCR primer used to
	13	23	1.6	738	19	AAV27046	Mouse neurogenin 1
	14	23	1.6	738	21	AAZ51977	Murine neurogenin-
	15	23	1.6	790	22	AAF27264	Chicken atonal hom
	16	23	1.6	1332	19	AAV42938	DNA encoding murin
	17	23	1.6	1333	18	AAV74894	Mouse neurogenic d
	18	23	1.6	1385	19	AAV27049	Mouse neurogenin 2
	19	23	1.6	1385	21	AAZ51980	Murine neurogenin-
	20	23	1.6	1385	22	AAF27269	Mouse neurogenin 2
	21	23	1.6	1412	22	AAF27255	Mouse atonal homol
	22	23	1.6	1412	22	AAF27273	Mouse atonal homol
	23	21	1.4	1268	18	AAV74891	Human neurogenic d
	24	21	1.4	1268	19	AAV42932	DNA encoding human
	25	21	1.4	1535	18	AAV74890	Human neurogenic d
	26	21	1.4	1535	19	AAV42931	DNA encoding human
	27	21	1.4	1550	21	AAAG2681	Human NeuroD2 gene
c	28	21	1.4	2776	22	AAAL04043	Human reproductive
c	29	21	1.4	2776	22	AAAL04045	Human reproductive
	30	21	1.4	2776	22	AAK68475	Human immune/haema
	31	21	1.4	2776	22	AAK68476	Human immune/haema
c	32	20	1.4	592	24	ABQ49522	Oligonucleotide fo
	33	20	1.4	592	24	ABQ49523	Oligonucleotide fo
c	34	19	1.3	50	17	AAT13245	Acetylcholine tran
c	35	19	1.3	50	17	AAT28933	nAChR beta2 subuni
c	36	19	1.3	510	24	ABQ91538	M. capsulatus gene
	37	19	1.3	615	24	ABQ91539	M. capsulatus gene
c	38	19	1.3	1260	24	ABQ91700	M. capsulatus gene
c	39	19	1.3	1263	21	AAC43008	Arabidopsis thalia
c	40	19	1.3	1698	21	AAC36256	Arabidopsis thalia
c	41	19	1.3	3424	22	AAAL06199	Human reproductive
c	42	19	1.3	3424	23	ABL98764	Human testicular a
c	43	19	1.3	3425	22	AAAL06200	Human reproductive
c	44	19	1.3	3425	23	ABL98765	Human testicular a
c	45	19	1.3	18733	22	AAK80682	Human immune/haema

ALIGNMENTS

RESULT 1

AAV42512

ID AAV42512 standard; cdNA; 1491 BP.

XX AAV42512;

AC AAV42512;

XX 05-OCT-1998 (first entry)

DT CDNA encoding a novel BHLH protein designated RELAX.

DE Basic helix-loop-helix; BHLH; RELAX; Rat Embryonic Longitudinal Axis;

KW control; gene expression; transcriptional activator; targeting;

KW protein expression; central nervous system; CNS; treatment;

KW nervous system disorder; ss.

XX Rattus sp.

OS Rattus sp.

XX Key

XX Location/Qualifiers

FT 459..1103

FT /*tag= a

FT /product= RELAX

XX WO9827206-A2.

XX 25-JUN-1998.

XX 19-DEC-1997; 97WO-FR02368.

XX 19-DEC-1996; 96FR-0015651.

XX (RHON) RHONE-POULENC RORER SA.

XX

XX WO9813491-A2.
 PN
 XX
 XX 02-APR-1998.
 PD
 XX
 XX 24-SEP-1997; 97WO-US17048.
 PF
 XX
 XX 17-SEP-1997; 97US-0932411.
 PR
 XX 27-SEP-1996; 96US-0722570.
 PR
 XX 12-NOV-1996; 96US-0030864.
 PR
 XX 19-DEC-1996; 96US-00772009.
 XX
 XX (CALY) CALIFORNIA INST OF TECHNOLOGY.
 PA
 XX Anderson DJ, Ma Q, Sommer L;
 PI
 XX WPI; 1998-230702/20.
 DR
 XX P-PSDB; AAW54947.
 DR
 XX
 XX Mouse neurogenins, useful in neurogenesis - and recombinant nucleic
 PT
 XX acids and proteins derived from rat and xenopus
 PT
 XX
 XX Disclosure; Fig 9; 106pp; English.
 PS
 XX The Mouse neurogenin 3 is one of several neurogenin proteins discussed
 CC
 XX in the present invention. The neurogenin nucleic acids can be expressed
 CC
 XX in a host cell, transformed using an expression vector, to produce
 CC
 XX recombinant proteins. The proteins and the antibodies raised against
 CC
 XX the proteins are useful in the study of neurogenesis.
 CC
 XX
 XX Sequence 804 BP; 171 A; 263 C; 225 G; 145 T; 0 other;
 SQ
 Query Match 6.3%; Score 92; DB 19; Length 804;
 Best Local Similarity 100.0%; Pred. No. 9.3e-35;
 Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 762 GCGCTGGATGCGCTGCGGTGCTCTGCCACCTTCCCGATGACGCCAACTTACAAG 821
 DB 463 GCGCTGGATGCGCTGCGGTGCTCTGCCACCTTCCCGATGACGCCAACTTACAAG 522
 QY 822 ATCGAGACCTGCGCTTCCGCCACAACTACAT 853
 DB 523 ATCGAGACCTGCGCTTCCGCCACAACTACAT 554
 RESULT 3
 AA251981
 ID AA251981 standard; DNA; 804 BP.
 XX
 XX AA251981;
 AC
 XX
 XX 04-JUL-2000 (first entry)
 DT
 XX
 XX Murine neurogenin-3 (NGN3) nucleic acid sequence.
 DE
 XX Neurogenin-3; NGN-3; non-neuronal cell; NNC; neurogenesis;
 KW Phox2a protein; neuronal subtype-specific marker; growth factor;
 KW neural differentiation; transplantation; neuronal dysfunction;
 KW optical nerve damage; auditory nerve damage, neurodegenerative disorder;
 KW neuroprotective; nontropic; anticonvulsant; antiParkinsonian; vulnary;
 KW cerebroprotective; immunosuppressant; antiinfectious; ds.
 XX
 XX Mus sp.
 OS
 XX
 XX Key Location/Qualifiers
 FH 160..804
 FT CDS /*tag= a
 FT /product= "Murine neurogenin-3 protein"
 FT
 XX
 XX WO200009676-A2.
 PN
 XX
 XX 24-FEB-2000.
 PD
 XX

PF 13-AUG-1999; 99WO-US18525.
 XX
 PR 14-AUG-1998; 98US-0096630.
 XX
 PA (CALY) CALIFORNIA INST OF TECHNOLOGY.
 XX
 XX Anderson DJ, Lo L;
 PI
 XX WPI; 2000-256250/22.
 DR
 XX P-PSDB; AAY70570.
 DR
 XX
 XX Inducing non-neuronal cells to differentiate into neurons and for
 PT
 XX non-neuronal cells to express a neuronal subtype-specific marker,
 PT
 XX comprising contacting the non-neuronal cells with a vector containing
 PT
 XX neurogenin nucleic acid -
 XX
 XX Claim 1; Fig 1J; 76pp; English.
 PS
 XX The patent discloses a method for inducing non-neuronal cells (NNC) to
 CC
 XX differentiate into neurons and for NNCs to express a neuronal subtype
 CC
 XX -specific marker. Transformed host cells are used as sources of neuronal
 CC
 XX and other growth factors; in culture for screening compounds that
 CC
 XX modulate neural differentiation or as sources of recombinantly produced
 CC
 XX neurogenins and Phox2a proteins for use in transplantation. The cells
 CC
 XX also have a variety of in vivo uses, e.g. for transplantation at sites of
 CC
 XX neuronal dysfunction e.g. patients with hearing or vision loss due to
 CC
 XX optical or auditory nerve damage, brain or spinal cord injuries, and
 CC
 XX neurodegenerative disorders e.g. Alzheimer's disease. The present
 CC
 XX sequence encodes murine neurogenin-3 (NGN-3), a transcription factor of
 CC
 XX NNCs that differentiate into neurons through the recombinant expression of a
 CC
 XX transcription factor that induces a core program of neurogenesis. Forced
 CC
 XX expression of murine NGN3 can elicit expression of at least some neuronal
 CC
 XX phenotypic markers even in NNCs.
 XX
 XX Sequence 804 BP; 171 A; 263 C; 225 G; 145 T; 0 other;
 SQ
 Query Match 6.3%; Score 92; DB 21; Length 804;
 Best Local Similarity 100.0%; Pred. No. 9.3e-35;
 Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 762 GCGCTGGATGCGCTGCGGTGCTCTGCCACCTTCCCGATGACGCCAACTTACAAG 821
 DB 463 GCGCTGGATGCGCTGCGGTGCTCTGCCACCTTCCCGATGACGCCAACTTACAAG 522
 QY 822 ATCGAGACCTGCGCTTCCGCCACAACTACAT 853
 DB 523 ATCGAGACCTGCGCTTCCGCCACAACTACAT 554
 RESULT 4
 AAF27266
 ID AAF27266 standard; cDNA; 861 BP.
 XX
 XX AAF27266;
 AC
 XX
 XX 24-APR-2001 (first entry)
 DT
 XX
 XX Mouse neurogenin 3 (ngn3) cDNA, SEQ ID NO:24.
 DE
 XX Atonal; homologue; orthologue; atonal-associated protein; deafness;
 KW hearing impairment; vestibular effect; balance disorder; osteoarthritis;
 KW cellular proliferation; cerebellar granule neuron; gene therapy;
 KW mechanoreceptive cell growth; auditory; osteopathic; cytostatic;
 KW transgenic animal; ss.
 XX
 XX Mus musculus.
 OS
 XX
 XX WO200073764-A2.
 PN
 XX
 XX 07-DEC-2000.
 PD
 XX
 XX 01-JUN-2000; 2000WO-US15410.
 PF
 XX

XX 01-JUN-2000; 2000WO-US15410.
 PF
 XX
 XX 01-JUN-1999; 99US-0137060.
 PR
 XX 19-JAN-2000; 2000US-0176993.
 PR
 XX
 XX (BAYU) BAYLOR COLLEGE MEDICINE.
 PA
 XX
 XX Zoghbi HY, Bellen H, Birmingham N, Hassan B, Ben-Arie N;
 PI
 XX
 XX WPI; 2001-032190/04.
 DR
 XX P-PSDB; AAB60350.
 DR
 XX
 XX Therapeutic use of atonal-associated nucleic acids or amino acids, or
 PT any of its homologs or orthologs, for the treatment of e.g. deafness,
 PT osteoarthritis and abnormal cell proliferation -
 XX
 XX
 XX Disclosure; Page : 142pp; English.
 PS
 XX
 XX The invention relates to the use of atonal-associated nucleic acid or
 CC amino acid sequence, or any of its homologues or orthologues as
 CC therapeutic agents for the treatment of deafness, partial hearing loss,
 CC vestibular effects due to damage or loss of inner hair cells,
 CC osteoarthritis and abnormal cell proliferation. The invention also
 CC encompasses methods of screening for compounds which affect the
 CC expression of an atonal-associated nucleic acid sequence in an animal,
 CC and a transgenic animal in which an allele of a native atonal-associated
 CC gene is replaced by a heterologous nucleic acid sequence, thus
 CC inactivating the atonal-associated allele. The nucleic acids or proteins
 CC may be used in a method of treating an animal for hearing impairment,
 CC joint disease, balance disorders, abnormal cell proliferation, or other
 CC disease related to loss of a functional atonal-associated nucleic acid or
 CC protein. They may particularly be used to treat an animal with a
 CC deficiency in cerebellar granule neurons or their precursors, and may
 CC also be used in promoting mechanoreceptive cell growth and generating
 CC hair cells. The present sequence represents an atonal-associated nucleic
 CC acid sequence referred to in the invention.
 CC Note: The present sequence is not shown in the specification, but
 CC was obtained from GenBank.
 XX
 XX Sequence 5567 BP; 1271 A; 1549 C; 1564 G; 1183 T; 0 other;
 SQ
 Query Match 6.3%; Score 92; DB 22; Length 5567;
 Best Local Similarity 100.0%; Pred. No. 8.5e-35;
 Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 762 GCGCTGGATGCGCTGCGGGTGCTCTGCCACCTTCCCGGATGACGCCAACTTACAAG 821
 Db
 5226 GCGCTGGATGCGCTGCGGGTGCTCTGCCACCTTCCCGGATGACGCCAACTTACAAG 5285
 QY 822 ATCGAGACCTGCGCTTGGCCCACTACAT 853
 Db 5286 ATCGAGACCTGCGCTTGGCCCACTACAT 5317
 RESULT 7
 ABN31392
 ID ABN31392 standard; DNA; 65 BP.
 XX
 XX AC
 XX ABN31392;
 XX
 XX
 XX 15-JUL-2002 (first entry)
 DT
 DE Rat spliced transcript detection oligonucleotide SEQ ID NO:4140.
 DE Human; mouse; rat; splice transcript; detection; RNA transcript;
 KW splice variant; transcriptome; oligonucleotide library; ss.
 XX
 XX Rattus norvegicus.
 OS
 XX WO200210449-A2.
 PN
 XX
 XX 07-FEB-2002.
 PD

XX 20-JUL-2001; 2001WO-IB01903.
 PF
 XX
 XX 28-JUL-2000; 2000US-221607P.
 PR
 XX 02-MAY-2001; 2001US-287724P.
 PR
 XX
 XX (COMP-) COMPUGEN INC.
 PA
 XX
 XX Shoshan A, Wasserman A, Mintz E, Mintz L, Faigler S;
 PI
 XX
 XX WPI; 2002-257383/30.
 DR
 XX
 XX
 XX New oligonucleotide libraries comprising oligonucleotides which
 PT selectively hybridize to mRNAs transcribed from a transcription unit of
 PT a genome, useful for detecting tissue-, pathology-, and
 PT developmental-specific genes
 XX
 XX Example 1; SEQ ID 4140; 47pp; English.
 PS
 XX
 XX The present invention describes oligonucleotide libraries for detecting
 CC messenger RNAs that populate a (sub-)transcriptome, where the
 CC (sub-)transcriptome comprises messenger RNAs transcribed from multiple
 CC transcription units that populate a genome. The library comprises
 CC several oligonucleotides, each capable of hybridizing selectively to a
 CC set of messenger RNAs transcribed from a given transcription unit of
 CC the genome, which encodes one or more messenger RNA splice variants.
 CC The oligonucleotide libraries are useful for detecting mRNAs from a
 CC biological sample, in expression profiling studies, in qualitatively or
 CC quantitatively characterizing the corresponding transcriptome, and in
 CC detecting RNA transcripts and splice variants of human or animal
 CC transcriptomes. The libraries may also be used as specialised mini
 CC libraries to detect transcripts of a sub-transcriptome under a
 CC particular biological or pathological state, and so allowing the
 CC detection of tissue- and pathology-specific genes such as those genes
 CC only expressed in specific tissue under a specific pathological
 CC condition; to detect developmental specific genes; and to detect RNA
 CC transcripts and splice variants of a transcriptome of a patient suffering
 CC from a particular disorder. ABN27253 to ABN59589 represent
 CC oligonucleotide sequences from rats, humans and mice, which are used in
 CC the exemplification of the present invention.
 CC N.B. The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 XX Sequence 65 BP; 11 A; 20 C; 14 G; 20 T; 0 other;
 SQ
 Query Match 4.5%; Score 65; DB 24; Length 65;
 Best Local Similarity 100.0%; Pred. No. 1.8e-21;
 Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1252 TTCCTTGCTGCTGCGTGCCACAAAGGACATTGCAGGCTGATCTCTCTTACCCCTCTC 1311
 Db 1 TTCCTTGCTGCTGCGTGCCACAAAGGACATTGCAGGCTGATCTCTCTTACCCCTCTC 60
 QY 1312 AGTGT 1316
 Db 61 AGTGT 65
 RESULT 8
 ABN57521
 ID ABN57521 standard; DNA; 65 BP.
 XX
 XX AC
 XX ABN57521;
 XX
 XX 15-JUL-2002 (first entry)
 DT
 DE Mouse spliced transcript detection oligonucleotide SEQ ID NO:30269.
 DE Human; mouse; rat; splice transcript; detection; RNA transcript;
 KW splice variant; transcriptome; oligonucleotide library; ss.
 XX
 XX Mus musculus.
 OS

```

XX PN WO200210449-A2.
XX PD 07-FEB-2002.
XX PF 20-JUL-2001; 2001WO-IB01903.
XX PR 28-JUL-2000; 2000US-221607P.
XX PR 02-MAY-2001; 2001US-287724P.
XX PA (COMP-) COMPUGEN INC.
XX PI Shoshan A, Wasserman A, Mintz E, Mintz L, Faigler S;
XX PI WPI; 2002-257383/30.
XX DR New oligonucleotide libraries comprising oligonucleotides which
XX PT selectively hybridize to mRNAs transcribed from a transcription unit of
XX PT a genome, useful for detecting tissue-, pathology-, and
XX PT developmental-specific genes
XX PS Example 1; SEQ ID 30269; 47pp; English.
XX CC The present invention describes oligonucleotide libraries for detecting
XX CC messenger RNAs that populate a (sub-)transcriptome, where the
XX CC (sub-)transcriptome comprises messenger RNAs transcribed from multiple
XX CC transcription units that populate a genome. The library comprises
XX CC several oligonucleotides, each capable of hybridizing selectively to a
XX CC set of messenger RNAs transcribed from a given transcription unit of
XX CC the genome, which encodes one or more messenger RNA splice variants.
XX CC The oligonucleotide libraries are useful for detecting mRNAs from a
XX CC biological sample, in expression profiling studies, in qualitatively or
XX CC quantitatively characterizing the corresponding transcriptome, and in
XX CC detecting RNA transcripts and splice variants of human or animal
XX CC transcriptomes. The libraries may also be used as specialised mini
XX CC libraries to detect transcripts of a sub-transcriptome under a
XX CC particular biological or pathological state, and so allowing the
XX CC detection of tissue- and pathology-specific genes such as those genes
XX CC only expressed in specific tissue under a specific pathological
XX CC condition; to detect developmental specific genes; and to detect RNA
XX CC transcripts and splice variants of a transcriptome of a patient suffering
XX CC from a particular disorder. ABN27253 to ABN59589 represent
XX CC oligonucleotide sequences from rats, humans and mice, which are used in
XX CC the exemplification of the present invention.
XX CC N.B. The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 65 BP; 16 A; 25 C; 11 G; 13 T; 0 other;

  Query Match      2.4%; Score 35; DB 24; Length 65;
  Best Local Similarity 100.0%; Pred. No. 8.8e-07;
  Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 547 TCAGTTCGAATTCACCCACCCACCTAGCCCTCTC 581
Db 1 TCAGTTCGAATTCACCCACCCACCTAGCCCTCTC 35

RESULT 9
AAC61089
ID AAC61089 standard; DNA: 5340 BP.
XX AC AAC61089;
XX DT 05-FEB-2001 (first entry)
XX DE Human neurogenin 3 (Ngn3) genomic DNA sequence.
XX KW Neurogenin 3; Ngn3; chromosome 10q22.1-22.2; cellular differentiation;
XX KW Islet cell precursor identification; diabetes mellitus; human; ds.
XX OS Homo sapiens.

```

```

XX Key Location/Qualifiers
XX CDS 3022..3666
XX FT /*tag- a "Ngn3"
XX FT /product- "Neurogenin 3"
XX FT /note- "Neurogenin 3"
XX WO2000059936-A1.
XX 12-OCT-2000.
XX 28-MAR-2000; 2000WO-US08436.
XX 06-APR-1999; 99US-0128180.
XX (REGC ) UNIV CALIFORNIA.
XX German MS, Lin J;
XX WPI; 2000-664989/64.
XX P-PSDB; AAY85617.
XX Novel human neurogenin 3 polypeptides and polynucleotides encoding
XX PT mellitus and to identify individuals at risk of diabetes -
XX PT Claim 6; Page 46-48; 54pp; English.
XX CC The human neurogenin 3 Ngn3 DNA sequence AAC61089 encodes the Ngn3
XX CC protein AAY85617. The Ngn3 gene is located at chromosome position
XX CC 10q22.1-22.2. The invention relates to the human Ngn3 nucleotide and
XX CC protein sequences, and includes an antibody recognising the Ngn3 protein.
XX CC Also included in the invention is a method for identifying an islet cell
XX CC precursor, the method involves analysing a cell for the expression of the
XX CC Ngn3 gene product, where detection of the product is indicative of an
XX CC islet cell precursor. The Ngn3 DNA sequence is useful as a diagnostic
XX CC reagent for detecting (in a subject) a predisposition to a defect in
XX CC pancreatic islet cell function or formation associated with a defect in
XX CC Ngn3 activity. The Ngn3 protein is useful for identifying beta-cell
XX CC precursor cells expressing Ngn3, and to alter cellular differentiation in
XX CC culture in vivo to produce new beta-cells to treat patients with diabetes
XX CC mellitus.
XX SQ Sequence 5340 BP; 1215 A; 1500 C; 1514 G; 1111 T; 0 other;

  Query Match      2.2%; Score 32; DB 21; Length 5340;
  Best Local Similarity 100.0%; Pred. No. 2.1e-05;
  Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 672 AAGAGCGAGTTGGCAGTCGACGACGACGACG 703
Db 3235 AAGAGCGAGTTGGCAGTCGACGACGACGACG 3266

RESULT 10
AAS33797
ID AAS33797 standard; CDNA: 428 BP.
XX AC AAS33797;
XX DT 17-DEC-2001 (first entry)
XX DE Human cDNA encoding a novel foetal antigen, SEQ ID No 321.
XX KW Human; foetal tissue antigen; ss; antinflammatory; neuroprotective;
XX KW immunomodulator; cardiovascular; cytostatic; nephrothropic;
XX KW cardiovascular; autoimmune disease; rheumatoid arthritis;
XX KW hyperproliferative disorder; breast neoplasm; cancer;
XX KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
XX KW cerebral ischaemia; angiogenesis; nervous system disorder;
XX KW Alzheimer's disease; infection; ocular disorder; corneal infection;
XX KW wound healing; epithelial cell proliferation; food additive.

```



```
XX Rosen CA, Barash SC, Ruben SM;
PI WPI; 2001-488782/53.
XX P-PSDB; AAU20977.
XX New polynucleotides and polypeptides for diagnosing, treating,
XX preventing or prognosing e.g. diseases or disorders of the nervous,
XX musculoskeletal, excretory, gastrointestinal, reproductive, and
XX respiratory systems
XX Claim 1; SEQ ID No 321; 642pp; English.
XX The invention relates to novel nucleic acids encoding novel human foetal
XX antigens. The nucleic acids and proteins are used to prevent, treat (e.g.
XX by gene therapy) or ameliorate a medical condition in e.g. humans, mice,
XX rabbits, goats, horses, cats, dogs, chickens or sheep. They
XX are also used in diagnosing a pathological condition or susceptibility
XX to a pathological condition. The antibodies to the antigens can also
XX be used in alleviating symptoms associated with the disorders and in
XX diagnostic immunoassays e.g. radioimmunoassays or enzyme linked
XX immunosorbent assays (ELISA). Disorders which are diagnosed or treated
XX include autoimmune diseases e.g. rheumatoid arthritis,
XX hyperproliferative disorders e.g. neoplasms of the breast or liver,
XX cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
XX e.g. cerebral ischaemia, angiodenesis, nervous system disorders e.g.
XX Alzheimer's disease, infections caused by bacteria, viruses and fungi
XX and ocular disorders e.g. corneal infection. The polypeptides can also
XX be used to aid wound healing and epithelial cell proliferation, to
XX prevent skin aging due to sunburn, to maintain organs before
XX transplantation, for supporting cell culture of primary tissues, to
XX regenerate tissues and in chemotaxis. The polypeptides can also be used
XX as a food additive or preservative to increase or decrease storage
XX capabilities, fat content, lipid, protein, carbohydrate, vitamins,
XX minerals, cofactors and other nutritional components. Numerous
XX examples of diseases and disorders treated by the nucleic acids and
XX proteins are given in the specification. The present sequence
XX
XX Query Match 2.0%; Score 29; DB 22; Length 428;
XX Best Local Similarity 100.0%; Pred. No. 0.0007;
XX Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 825 GAGACCTGGCGTTCGCCACACTACAT 853
XX |||||
XX Db 1 GAGACCTGGCGTTCGCCACACTACAT 29
XX
XX RESULT 11
XX ABT03700/c
XX ID ABT03700 standard; DNA; 26 BP.
XX AC ABT03700;
XX AC ABT03700;
XX DT 13-SEP-2002 (first entry)
XX DE Human Neurogenin-3 gene PCR primer SEQ ID NO: 221.
XX KW Human; cancer; neoplastic disease; tumour specific marker; cytostatic;
XX KW transcription factor; PCR; primer; ss.
XX OS Homo sapiens.
XX PN WO200240716-A2.
XX XX
XX PD 23-MAY-2002.
XX PF 13-NOV-2001; 2001WO-US43461.
XX PR 16-NOV-2000; 2000US-249508P.
XX XX
XX PA (CEMI-) CEMINES LLC.
XX XX
XX PI Palm K;
```

CC presence of an intact E box (CANNTG), particularly for targeting
 CC expression of proteins to the central nervous system (CNS). The
 CC nucleic acid sequence can be used to treat nervous system disorders,
 CC and antisense sequences can be used to control mRNA transcription.
 XX
 SQ Sequence 25 BP; 5 A; 9 C; 6 G; 5 T; 0 other;

Query Match 1.7%; Score 25; DB 19; Length 25;
 Best Local Similarity 100.0%; Pred. No. 0.073; 0; Indels 0; Gaps 0;
 Matches 25; Conservative 0; Mismatches 0;

QY 750 AACCTTAACCTCCGCGTGGATGCGC 774
 ||||||||||||||||||||||||||||
 DB 1 AACCTTAACCTCCGCGTGGATGCGC 25

RESULT 13
 AAV27046
 ID AAV27046 standard; cDNA; 738 BP.
 XX
 AC AAV27046;
 XX
 DT 17-SEP-1998 (first entry)
 XX
 DE Mouse neurogenin 1 gene.
 XX
 KW ss; Mouse; neurogenin; expression vector; recombinant protein;
 KW antibody; neurogenesis.
 XX
 OS Mus sp.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..735
 FT /*tag= a
 FT /product= "Mouse neurogenin 1"
 XX
 PN WO9813491-A2.
 XX
 PD 02-APR-1998.
 XX
 PF 24-SEP-1997; 97WO-US17048.
 XX
 PR 17-SEP-1997; 97US-0932411.
 PR 27-SEP-1996; 96US-0722570.
 PR 12-NOV-1996; 96US-0030864.
 PR 19-DEC-1996; 96US-0772009.
 XX
 PA (CALY) CALIFORNIA INST OF TECHNOLOGY.
 XX
 PI Anderson DJ, Ma Q, Sommer L;
 XX
 DR WPI; 1998-230702/20.
 DR P-PSDB; AAW54944.
 XX
 PT Mouse neurogenins, useful in neurogenesis - and recombinant nucleic
 PT acids and proteins derived from rat and Xenopus
 XX
 PS Claim 5; Fig 4; 106pp; English.
 XX
 CC The mouse neurogenin 1 is one of several neurogenin proteins discussed in
 CC the present invention. The neurogenin nucleic acids can be expressed in
 CC a host cell, transformed using an expression vector, to produce
 CC recombinant proteins. The proteins and the antibodies raised against
 CC the proteins are useful in the study of neurogenesis.
 XX
 SQ Sequence 738 BP; 121 A; 283 C; 205 G; 129 T; 0 other;

Query Match 1.6%; Score 23; DB 19; Length 738;
 Best Local Similarity 100.0%; Pred. No. 0.59;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 732 GAGCGCAACCGCATGCACCACT 754
 ||||||||||||||||||||||||

DB 304 GAGCGCAACCGCATGCACCACT 326

RESULT 14
 AAZ51977
 ID AAZ51977 standard; DNA; 738 BP.
 XX
 AC AAZ51977;
 XX
 DT 04-JUL-2000 (first entry)
 XX
 DE Murine neurogenin-1 (NGN1) nucleic acid sequence.
 XX
 KW Neurogenin-1; NGN-1; non-neuronal cell; NNC; neurogenesis;
 KW Phox2a protein; neuronal subtype-specific marker; growth factor;
 KW neural differentiation; transplantation; neuronal dysfunction; disorder;
 KW optical nerve damage; auditory nerve damage; neurodegenerative disorder;
 KW neuroprotective; nontropic; anticonvulsant; antiparkinsonian; vulnerary;
 KW cerebroprotective; immunosuppressant; antineoplastic; ss.
 XX
 OS Mus sp.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..735
 FT /*tag= a
 FT /product= "Murine neurogenin-1 protein"
 XX
 PN WO200009676-A2.
 XX
 PD 24-FEB-2000.
 XX
 PF 13-AUG-1999; 99WO-US18525.
 XX
 PR 14-AUG-1998; 98US-0096630.
 XX
 PA (CALY) CALIFORNIA INST OF TECHNOLOGY.
 XX
 PI Anderson DJ, Lo L;
 XX
 DR WPI; 2000-256250/22.
 DR P-PSDB; AAY70566.
 XX
 PT Inducing non-neuronal cells to differentiate into neurons and for
 PT non-neuronal cells to express a neuronal subtype-specific marker,
 PT comprising contacting the non-neuronal cells with a vector containing
 PT neurogenin nucleic acid -
 XX
 PS Claim 1; Fig 1C; 76pp; English.
 XX
 CC The patent discloses a method for inducing non-neuronal cells (NNC) to
 CC differentiate into neurons and for NNCs to express a neuronal subtype
 CC -specific marker. Transformed host cells are used as sources of neuronal
 CC and other growth factors; in culture for screening compounds that
 CC modulate neural differentiation or as sources of recombinantly produced
 CC neurogenins and Phox2a proteins for use in transplantation. The cells
 CC also have a variety of in vivo uses, e.g. for transplantation at sites of
 CC neuronal dysfunction e.g. patients with hearing or vision loss due to
 CC optical or auditory nerve damage, brain or spinal cord injuries, and
 CC neurodegenerative disorders e.g. Alzheimer's disease. The present
 CC sequence encodes murine neurogenin-1 (NGN-1), a transcription factor.
 CC NNCs differentiate into neurons through the recombinant expression of a
 CC transcription factor that induces a core program of neurogenesis. Forced
 CC expression of murine NGN1 can elicit expression of at least some neuronal
 CC phenotypic markers even in NNCs. This can be used in autografting.
 XX
 SQ Sequence 738 BP; 121 A; 283 C; 205 G; 129 T; 0 other;

Query Match 1.6%; Score 23; DB 21; Length 738;
 Best Local Similarity 100.0%; Pred. No. 0.59;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 732 GAGCGCAACCGCATGCACCACT 754
 ||||||||||||||||||||||||

Search completed: April 8, 2003, 21:20:26
Job time : 404 secs

Db 304 GAGCGCAACCGCATGCACAACT 326

RESULT 15
AAF27264
ID AAF27264 standard; CDNA; 790 BP.
XX
AC AAF27264;
XX
DT 24-APR-2001 (first entry)
XX
DE Chicken atonal homologue ngn2/ath4a cdna, SEQ ID NO:20.
XX
KW Atonal; homologue; orthologue; atonal-associated protein; deafness;
KW hearing impairment; vestibular effect; balance disorder; osteoarthritis;
KW cellular proliferation; cerebellar granule neuron; gene therapy;
KW mechanoreceptive cell growth; auditory; osteopathic; cytostatic;
KW transgenic animal; ss.
XX
OS Gallus gallus.
XX
PN WO200073764-A2.
XX
PD 07-DEC-2000.
XX
PF 01-JUN-2000; 2000WO-US15410.
XX
PR 01-JUN-1999; 99US-0137060.
XX
PR 19-JAN-2000; 2000US-0176993.
XX
XX (BAYU) BAYLOR COLLEGE MEDICINE.
XX
PA Zoghbi HY, Bellen H, Birmingham N, Hassan B, Ben-Arie N;
XX
PI WPI; 2001-032190/04.
XX
DR P-PSDB; AAB60357.
XX
PT Therapeutic use of atonal-associated nucleic acids or amino acids, or
PT any of its homologs or orthologs, for the treatment of e.g. deafness,
PT osteoarthritis and abnormal cell proliferation -
XX
PS Disclosure; Page -: 142pp; English.
XX
CC The invention relates to the use of atonal-associated nucleic acid or
CC amino acid sequence, or any of its homologues or orthologues as
CC therapeutic agents for the treatment of deafness, partial hearing loss,
CC vestibular effects due to damage or loss of inner hair cells,
CC osteoarthritis and abnormal cell proliferation. The invention also
CC encompasses methods of screening for compounds which affect the
CC expression of an atonal-associated nucleic acid sequence in an animal,
CC and a transgenic animal in which an allele of a native atonal-associated
CC gene is replaced by a heterologous nucleic acid sequence, thus
CC inactivating the atonal-associated allele. The nucleic acids or proteins
CC may be used in a method of treating an animal for hearing impairment,
CC joint disease, balance disorders, abnormal cell proliferation, or other
CC disease related to loss of a functional atonal-associated nucleic acid or
CC protein. They may particularly be used to treat an animal with a
CC deficiency in cerebellar granule neurons or their precursors, and may
CC also be used in promoting mechanoreceptive cell growth and generating
CC hair cells. The present sequence represents an atonal-associated nucleic
CC acid sequence referred to in the invention.
CC Note: The present sequence is not shown in the specification, but
CC was obtained from GenBank.
XX
SQ Sequence 790 BP; 91 A; 351 C; 283 G; 65 T; 0 other;

Query Match 1.6%; Score 23; DB 22; Length 790;
Best Local Similarity 100.0%; Pred. No. 0.59;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 732 GAGCGCAACCGCATGCACAACT 754
|||||
Db 374 GAGCGCAACCGCATGCACAACT 396
|||||

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 8, 2003, 21:09:53 ; Search time 164 Seconds
(without alignments)
7808.933 Million cell updates/sec

Title: US-09-595-947C-1
Perfect score: 1460
Sequence: 1 gcaggtagcagagagagcag.....agagtgcacttaaccagtgt 1460

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 593429 seqs, 438583890 residues

Word size : 15

Total number of hits satisfying chosen parameters: 1316

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PublishedApplications_NA.*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq.*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	92	6.3	645	9 US-10-004-717-4	Sequence 4, Appli
2	92	6.3	861	9 US-10-004-717-24	Sequence 24, Appl
3	92	6.3	1861	10 US-09-817-360-3	Sequence 3, Appli
4	32	2.2	5340	10 US-09-817-360-1	Sequence 1, Appli
5	23	1.6	738	7 US-08-722-570-13	Sequence 13, Appl
6	23	1.6	790	9 US-10-004-717-20	Sequence 20, Appl
7	23	1.6	1385	9 US-10-004-717-30	Sequence 30, Appl
8	23	1.6	1412	9 US-10-004-717-6	Sequence 6, Appli
9	23	1.6	1412	9 US-10-004-717-37	Sequence 37, Appl
10	20	1.4	352	10 US-09-728-445-108	Sequence 108, App
11	19	1.3	500	10 US-09-783-590-2157	Sequence 2157, Ap
12	19	1.3	1263	9 US-09-938-842A-1036	Sequence 1036, Ap
13	18	1.2	4797	10 US-09-751-797-25	Sequence 25, Appl
14	18	1.2	20272	10 US-09-908-711-145	Sequence 145, App
15	18	1.2	25619	10 US-09-908-711-143	Sequence 143, App
16	18	1.2	25619	10 US-09-764-898-302	Sequence 302, App
17	18	1.2	31728	9 US-10-114-170-64	Sequence 64, Appl
18	18	1.2	402850	9 US-09-844-653-5	Sequence 5, Appli
19	17	1.2	31	10 US-09-817-360-17	Sequence 17, Appl

20	17	1.2	232	10 US-09-960-352-14228	Sequence 14228, A
21	17	1.2	285	10 US-09-960-352-6887	Sequence 6887, Ap
22	17	1.2	285	10 US-09-960-352-8778	Sequence 8778, Ap
23	17	1.2	307	10 US-09-783-590-12149	Sequence 12149, A
24	17	1.2	454	9 US-10-040-739-1094	Sequence 1094, Ap
25	17	1.2	454	10 US-09-864-761-257	Sequence 257, App
26	17	1.2	469	10 US-09-864-761-5374	Sequence 5374, Ap
27	17	1.2	471	10 US-09-864-761-11	Sequence 11, Appl
28	17	1.2	473	10 US-09-864-761-22148	Sequence 22148, A
29	17	1.2	474	10 US-09-864-761-353	Sequence 353, App
30	17	1.2	474	10 US-09-864-761-10932	Sequence 10932, A
31	17	1.2	487	10 US-09-864-761-3320	Sequence 3320, Ap
32	17	1.2	495	10 US-09-864-761-5673	Sequence 5673, Ap
33	17	1.2	499	10 US-09-783-590-4421	Sequence 4421, Ap
34	17	1.2	594	10 US-09-864-761-7938	Sequence 7938, Ap
35	17	1.2	756	10 US-09-910-943-147	Sequence 147, App
36	17	1.2	1074	9 US-10-004-717-18	Sequence 18, Appl
37	17	1.2	1238	9 US-09-991-496-75	Sequence 75, Appl
38	17	1.2	1238	10 US-09-874-923-75	Sequence 75, Appl
39	17	1.2	1362	9 US-09-970-966-208	Sequence 208, App
40	17	1.2	1362	10 US-09-825-294-208	Sequence 208, App
41	17	1.2	1422	10 US-09-815-242-4334	Sequence 4334, Ap
42	17	1.2	1425	10 US-09-815-242-8558	Sequence 8558, Ap
43	17	1.2	1669	10 US-09-880-107-2159	Sequence 2159, Ap
44	17	1.2	1747	9 US-09-764-868-1334	Sequence 1334, Ap
45	17	1.2	1752	9 US-09-764-868-137	Sequence 137, App

ALIGNMENTS

RESULT 1
US-10-004-717-4
; Sequence 4, Application US/10004717
; Publication No. US20020192665A1
; GENERAL INFORMATION:
; APPLICANT: ZOGHBI, HUDAY.
; APPLICANT: YANG, QI
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN ATONAL ASSOCIATED SEQUENCE FOR DEAFNESS,
; TITLE OF INVENTION: OSTEOARTHRITIS, AND ABNORMAL CELL PROLIFERATION
; FILE REFERENCE: P01899054
; CURRENT APPLICATION NUMBER: US/10/004,717
; CURRENT FILING DATE: 2002-08-16
; PRIOR APPLICATION NUMBER: 09/585,645
; PRIOR FILING DATE: 2000-06-01
; PRIOR APPLICATION NUMBER: 60/176,993
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: 60/137,060
; PRIOR FILING DATE: 1999-06-01
; NUMBER OF SEQ ID NOS: 69
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 645
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-004-717-4

Query Match 6.3%; Score 92; DB 9; Length 645;
Best Local Similarity 100.0%; Pred. No. 5.5e-39;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 762 GCGCTGGATGCGCTCGGGGTGCTCTGCCACCTTCCCGGATGAGCCAACTTCAAG 821
|||||
Db 304 GCGCTGGATGCGCTCGGGGTGCTCTGCCACCTTCCCGGATGAGCCAACTTCAAG 363
QY 822 ATCGAGACCTCGCTCGGCCACACACTACAT 853
|||||
Db 364 ATCGAGACCTCGCTCGGCCACACACTACAT 395

RESULT 2
US-10-004-717-24


```
; Sequence 24, Application US/10004717
; Publication No. US20020192665A1
; GENERAL INFORMATION:
; APPLICANT: ZOGHBI, HUDA Y.
; APPLICANT: YANG, QI
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN
; TITLE OF INVENTION: ATONAL ASSOCIATED SEQUENCE FOR DEAFNESS.
; TITLE OF INVENTION: OSTEOARTHRITIS, AND ABNORMAL CELL PROLIFERATION
; FILE REFERENCE: P01899054
; CURRENT APPLICATION NUMBER: US/10/004,717
; CURRENT FILING DATE: 2002-08-16
; PRIOR APPLICATION NUMBER: 09/585,645
; PRIOR FILING DATE: 2000-06-01
; PRIOR APPLICATION NUMBER: 60/176,993
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: 60/137,060
; PRIOR FILING DATE: 1999-06-01
; NUMBER OF SEQ ID NOS: 69
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 24
; LENGTH: 861
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-004-717-24

Query Match 6.3%; Score 92; DB 9; Length 861;
Best Local Similarity 100.0%; Pred. No. 5.4e-39;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 762 GCCTGTGATCGCTGCGGTGCTCTGCCACCTTCCGGATGACGCCAACTTACAAAG 821
Db 463 GCCTGTGATCGCTGCGGTGCTCTGCCACCTTCCGGATGACGCCAACTTACAAAG 522

QY 822 ATCGAGACCTGCGCTTCGCCCAACTACAT 853
Db 523 ATCGAGACCTGCGCTTCGCCCAACTACAT 554

RESULT 3
US-09-817-360-3
; Sequence 3, Application US/09817360
; Patent No. US20020015696A1
; GENERAL INFORMATION:
; APPLICANT: German, Michael S.
; APPLICANT: Lin, Joseph
; TITLE OF INVENTION: PRODUCTION OF PANCREATIC ISLET CELLS
; TITLE OF INVENTION: AND DELIVERY OF INSULIN
; FILE REFERENCE: UCSF-129CIP
; CURRENT APPLICATION NUMBER: US/09/817,360
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: 09/535,145
; PRIOR FILING DATE: 2000-03-24
; PRIOR APPLICATION NUMBER: 60/128,180
; PRIOR FILING DATE: 1999-04-06
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1861
; TYPE: DNA
; ORGANISM: Mus musculus
US-09-817-360-3

Query Match 6.3%; Score 92; DB 10; Length 1861;
Best Local Similarity 100.0%; Pred. No. 5e-39;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 762 GCCTGTGATCGCTGCGGTGCTCTGCCACCTTCCGGATGACGCCAACTTACAAAG 821
Db 1396 GCCTGTGATCGCTGCGGTGCTCTGCCACCTTCCGGATGACGCCAACTTACAAAG 1455

QY 822 ATCGAGACCTGCGCTTCGCCCAACTACAT 853
Db 1456 ATCGAGACCTGCGCTTCGCCCAACTACAT 1487
```

```
RESULT 4
US-09-817-360-1
; Sequence 1, Application US/09817360
; Patent No. US20020015696A1
; GENERAL INFORMATION:
; APPLICANT: German, Michael S.
; APPLICANT: Lin, Joseph
; TITLE OF INVENTION: PRODUCTION OF PANCREATIC ISLET CELLS
; TITLE OF INVENTION: AND DELIVERY OF INSULIN
; FILE REFERENCE: UCSF-129CIP
; CURRENT APPLICATION NUMBER: US/09/817,360
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: 09/535,145
; PRIOR FILING DATE: 2000-03-24
; PRIOR APPLICATION NUMBER: 60/128,180
; PRIOR FILING DATE: 1999-04-06
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 5340
; TYPE: DNA
; ORGANISM: Homo Sapiens
US-09-817-360-1

Query Match 2.2%; Score 32; DB 10; Length 5340;
Best Local Similarity 100.0%; Pred. No. 4.9e-07;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 672 AAGAGCGAGTTGGCACTGACGACGACGACG 703
Db 3235 AAGAGCGAGTTGGCACTGACGACGACGACG 3266

RESULT 5
US-08-722-570-13
; Sequence 13, Application US/08722570
; Publication No. US20030044887A1
; GENERAL INFORMATION:
; APPLICANT: Anderson, David J.
; APPLICANT: Ma, Qifu
; TITLE OF INVENTION: NEUROGENIN
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/722,570
; FILING DATE: 27-SEP-1996
; CLASSIFICATION: 5365
; ATTORNEY/AGENT INFORMATION:
; NAME: Silva, Robin M.
; REGISTRATION NUMBER: 38,304
; REFERENCE/DOCKET NUMBER: A-63902/RFT/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 738 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
```

; TOPOLOGY: unknown
; MOLECULE TYPE: DNA
US-08-722-570-13

Query Match 1.6%; Score 23; DB 7; Length 738;
Best Local Similarity 100.0%; Pred. No. 0.038; 0; Indels 0; Gaps 0;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 732 GAGCGCAACCGCATGCACACCT 754
|||||
DB 304 GAGCGCAACCGCATGCACACCT 326

RESULT 6

US-10-004-717-20
; Sequence 20, Application US/10004717
; Publication No. US20020192665A1

; GENERAL INFORMATION:

; APPLICANT: ZOGHBI, HUDA Y.

; APPLICANT: YANG, QI

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN
; TITLE OF INVENTION: ATONAL ASSOCIATED SEQUENCE FOR DEAFNESS,

; FILE REFERENCE: P01899US4

; CURRENT APPLICATION NUMBER: US/10/004,717

; PRIOR FILING DATE: 2002-08-16

; PRIOR APPLICATION NUMBER: 09/585,645

; PRIOR FILING DATE: 2000-06-01

; PRIOR APPLICATION NUMBER: 60/176,993

; PRIOR FILING DATE: 2000-01-19

; PRIOR APPLICATION NUMBER: 60/137,060

; PRIOR FILING DATE: 1999-06-01

; NUMBER OF SEQ ID NOS: 69

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 20

; LENGTH: 790

; TYPE: DNA

; ORGANISM: chicken

US-10-004-717-20

Query Match

Best Local Similarity 1.6%; Score 23; DB 9; Length 790;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 732 GAGCGCAACCGCATGCACACCT 754
|||||
DB 374 GAGCGCAACCGCATGCACACCT 396

RESULT 7

US-10-004-717-30
; Sequence 30, Application US/10004717
; Publication No. US20020192665A1

; GENERAL INFORMATION:

; APPLICANT: ZOGHBI, HUDA Y.

; APPLICANT: YANG, QI

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN
; TITLE OF INVENTION: ATONAL ASSOCIATED SEQUENCE FOR DEAFNESS,

; FILE REFERENCE: P01899US4

; CURRENT APPLICATION NUMBER: US/10/004,717

; PRIOR FILING DATE: 2002-08-16

; PRIOR APPLICATION NUMBER: 09/585,645

; PRIOR FILING DATE: 2000-06-01

; PRIOR APPLICATION NUMBER: 60/176,993

; PRIOR FILING DATE: 2000-01-19

; PRIOR APPLICATION NUMBER: 60/137,060

; PRIOR FILING DATE: 1999-06-01

; NUMBER OF SEQ ID NOS: 69

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 30

; LENGTH: 1385

; TYPE: DNA

; ORGANISM: Mus musculus

; ORGANISM: Mus musculus
US-10-004-717-30

Query Match 1.6%; Score 23; DB 9; Length 1385;
Best Local Similarity 100.0%; Pred. No. 0.036;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 732 GAGCGCAACCGCATGCACACCT 754
|||||
DB 742 GAGCGCAACCGCATGCACACCT 764

RESULT 8

US-10-004-717-6
; Sequence 6, Application US/10004717
; Publication No. US20020192665A1

; GENERAL INFORMATION:

; APPLICANT: ZOGHBI, HUDA Y.

; APPLICANT: YANG, QI

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN
; TITLE OF INVENTION: ATONAL ASSOCIATED SEQUENCE FOR DEAFNESS,

; FILE REFERENCE: P01899US4

; CURRENT APPLICATION NUMBER: US/10/004,717

; PRIOR FILING DATE: 2002-08-16

; PRIOR APPLICATION NUMBER: 09/585,645

; PRIOR FILING DATE: 2000-06-01

; PRIOR APPLICATION NUMBER: 60/176,993

; PRIOR FILING DATE: 2000-01-19

; PRIOR APPLICATION NUMBER: 60/137,060

; PRIOR FILING DATE: 1999-06-01

; NUMBER OF SEQ ID NOS: 69

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 6

; LENGTH: 1412

; TYPE: DNA

; ORGANISM: Mus musculus

US-10-004-717-6

Query Match

Best Local Similarity 1.6%; Score 23; DB 9; Length 1412;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 732 GAGCGCAACCGCATGCACACCT 754
|||||
DB 445 GAGCGCAACCGCATGCACACCT 467

RESULT 9

US-10-004-717-37
; Sequence 37, Application US/10004717
; Publication No. US20020192665A1

; GENERAL INFORMATION:

; APPLICANT: ZOGHBI, HUDA Y.

; APPLICANT: YANG, QI

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN
; TITLE OF INVENTION: ATONAL ASSOCIATED SEQUENCE FOR DEAFNESS,

; FILE REFERENCE: P01899US4

; CURRENT APPLICATION NUMBER: US/10/004,717

; PRIOR FILING DATE: 2002-08-16

; PRIOR APPLICATION NUMBER: 09/585,645

; PRIOR FILING DATE: 2000-06-01

; PRIOR APPLICATION NUMBER: 60/176,993

; PRIOR FILING DATE: 2000-01-19

; PRIOR APPLICATION NUMBER: 60/137,060

; PRIOR FILING DATE: 1999-06-01

; NUMBER OF SEQ ID NOS: 69

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 37

; LENGTH: 1412

; TYPE: DNA

; ORGANISM: Mus musculus

US-10-004-717-37

Query Match 1.48; Score 23; DB 9; Length 1412;
Best Local Similarity 100.0%; Pred. No. 0.036;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 732 GAGCGCAACCGCATGCACAACT 754
Db 445 GAGCGCAACCGCATGCACAACT 467

RESULT 10

US-09-728-445-108/c
; Sequence 108, Application US/09728445
; Patent No. US20020102543A1
; GENERAL INFORMATION:
; APPLICANT: Friedrich, Glenn
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: No. US20020102543A1 Mutated Mammalian Cells and
; TITLE OF INVENTION: Animals
; FILE REFERENCE: LEX-0102-USA
; CURRENT APPLICATION NUMBER: US/09/728,445
; CURRENT FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/168,358
; PRIOR FILING DATE: 1999-12-01
; NUMBER OF SEQ ID NOS: 891
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 108
; LENGTH: 352
; TYPE: DNA
; ORGANISM: Mus musculus
US-09-728-445-108

Query Match 1.48; Score 20; DB 10; Length 352;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1010 CACAGCCTCATGGAGGAGT 1029
Db 209 CACAGCCTCATGGAGGAGT 190

RESULT 11

US-09-783-590-2157/c
; Sequence 2157, Application US/09783590
; Patent No. US2002010850A1
; GENERAL INFORMATION:
; APPLICANT: Dillon, Patrick J.
; APPLICANT: Haseltine, William A.
; APPLICANT: Li, Haodong
; APPLICANT: Rosen, Craig A.
; APPLICANT: Ruben, Steven M.
; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2
; FILE REFERENCE: PO#16.2C1
; CURRENT APPLICATION NUMBER: US/09/783,590
; CURRENT FILING DATE: 2000-02-15
; PRIOR APPLICATION NUMBER: 08/420,856
; PRIOR FILING DATE: 1995-04-12
; PRIOR APPLICATION NUMBER: 08/346,731
; PRIOR FILING DATE: 1994-11-21
; NUMBER OF SEQ ID NOS: 12485
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2157
; LENGTH: 500
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (84)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (207)

; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (215)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (324)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (354)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (360)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (368)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (369)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (379)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (393)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (411)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (418)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (439)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (440)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (441)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (451)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-783-590-2157

Query Match 1.38; Score 19; DB 10; Length 500;
Best Local Similarity 100.0%; Pred. No. 5.4;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1319 CCACCTCAAACTCCGCTC 1337
Db 23 CCACCTCAAACTCCGCTC 5

RESULT 12

US-09-938-842A-1036/c
; Sequence 1036, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE REFERENCE: SCRIPI300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111

;; PRIOR FILING DATE: 2001-06-22
;; NUMBER OF SEQ ID NOS: 5379
;; SEQ ID NO 1036
;; LENGTH: 1263
;; TYPE: DNA
;; ORGANISM: Arabidopsis thaliana
US-09-938-842A-1036

Query Match 1.3%; Score 19; DB 9; Length 1263;
Best Local Similarity 100.0%; Pred. No. 4.9;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 658 GCGCAACAGGCCCAAGAG 676
|||||
DB 599 GCGCAACAGGCCCAAGAG 581

RESULT 13
US-09-751-797-25
;; Sequence 25, Application US/09751797
;; Patent No. US20010024652A1
;; GENERAL INFORMATION:
;; APPLICANT: Dumoutier, Laure
;; APPLICANT: Loubet, Jamila
;; APPLICANT: Renaud, Jean-Christophe
;; TITLE OF INVENTION: Isolated Nucleic Acid Molecules which Encode T Cell Inducible Factors
;; TITLE OF INVENTION: (Tifs) The Proteins Encoded, and Uses Thereof
;; FILE REFERENCE: LUD 5543.2
;; CURRENT APPLICATION NUMBER: US/09/751,797
;; CURRENT FILING DATE: 2000-12-29
;; PRIOR APPLICATION NUMBER: 09/419,568
;; PRIOR FILING DATE: 1999-10-18
;; PRIOR APPLICATION NUMBER: US09/178,973
;; PRIOR FILING DATE: 1998-10-26
;; NUMBER OF SEQ ID NOS: 29
;; SEQ ID NO 25
;; LENGTH: 4797
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
US-09-751-797-25

Query Match 1.2%; Score 18; DB 10; Length 4797;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1152 GTCAGAGCTGCTGAAAT 1169
|||||
DB 3460 GTCAGAGCTGCTGAAAT 3477

RESULT 14
US-09-908-711-145/c
;; Sequence 145, Application US/09908711
;; Patent No. US20020045230A1
;; GENERAL INFORMATION:
;; APPLICANT: Rosen et al.
;; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;; FILE REFERENCE: PA128
;; CURRENT APPLICATION NUMBER: US/09/908,711
;; CURRENT FILING DATE: 2001-07-20
;; PRIOR APPLICATION NUMBER: US01/01360
;; PRIOR FILING DATE: 2001-01-17
;; PRIOR APPLICATION NUMBER: 09/764,867
;; PRIOR FILING DATE: 2001-01-17
;; PRIOR APPLICATION NUMBER: US01/01344
;; PRIOR FILING DATE: 2001-01-17
;; PRIOR APPLICATION NUMBER: 09/764,892
;; PRIOR FILING DATE: 2001-01-17
;; PRIOR APPLICATION NUMBER: US01/01345
;; PRIOR FILING DATE: 2001-01-17
;; PRIOR APPLICATION NUMBER: 09/764,888
;; PRIOR FILING DATE: 2001-01-17

;; PRIOR APPLICATION NUMBER: US01/01329
;; PRIOR FILING DATE: 2001-01-17
;; PRIOR APPLICATION NUMBER: 09/764,905
;; PRIOR FILING DATE: 2001-01-17
;; PRIOR APPLICATION NUMBER: US01/01354
;; PRIOR FILING DATE: 2001-01-17
;; PRIOR APPLICATION NUMBER: 09/764,891
;; PRIOR FILING DATE: 2001-01-17
;; PRIOR APPLICATION NUMBER: US01/01339
;; PRIOR FILING DATE: 2001-01-17
;; PRIOR APPLICATION NUMBER: 09/764,869
;; PRIOR FILING DATE: 2001-01-17
;; PRIOR APPLICATION NUMBER: US01/01340
;; PRIOR FILING DATE: 2001-01-17
;; PRIOR APPLICATION NUMBER: 09/764,874
;; PRIOR FILING DATE: 2001-01-17
;; PRIOR APPLICATION NUMBER: US01/01334
;; PRIOR FILING DATE: 2001-01-17
;; PRIOR APPLICATION NUMBER: 09/764,898
;; PRIOR FILING DATE: 2001-01-17
;; PRIOR APPLICATION NUMBER: US01/01320
;; PRIOR FILING DATE: 2001-01-17
;; PRIOR APPLICATION NUMBER: 09/764,853
;; PRIOR FILING DATE: 2001-01-17
;; PRIOR APPLICATION NUMBER: US01/01349
;; PRIOR FILING DATE: 2001-01-17
;; PRIOR APPLICATION NUMBER: 09/764,902
;; PRIOR FILING DATE: 2001-01-17
;; PRIOR APPLICATION NUMBER: US01/01239
;; PRIOR FILING DATE: 2001-01-17
;; PRIOR APPLICATION NUMBER: 09/764,870
;; PRIOR FILING DATE: 2001-01-17
;; PRIOR APPLICATION NUMBER: US01/01348
;; PRIOR FILING DATE: 2001-01-17
;; PRIOR APPLICATION NUMBER: 09/764,882
;; PRIOR FILING DATE: 2001-01-17
;; PRIOR APPLICATION NUMBER: US01/01347
;; PRIOR FILING DATE: 2001-01-17
;; PRIOR APPLICATION NUMBER: 09/764,896
;; PRIOR FILING DATE: 2001-01-17
;; PRIOR APPLICATION NUMBER: US01/01307
;; PRIOR FILING DATE: 2001-01-17
;; PRIOR APPLICATION NUMBER: 09/764,864
;; PRIOR FILING DATE: 2001-01-17
;; PRIOR APPLICATION NUMBER: US01/01341
;; PRIOR FILING DATE: 2001-01-17
;; PRIOR APPLICATION NUMBER: 09/764,856
;; PRIOR FILING DATE: 2001-01-17
;; PRIOR APPLICATION NUMBER: US01/01336
;; PRIOR FILING DATE: 2001-01-17
;; PRIOR APPLICATION NUMBER: 09/764,868
;; PRIOR FILING DATE: 2001-01-17
;; PRIOR APPLICATION NUMBER: US01/01312
;; PRIOR FILING DATE: 2001-01-17
;; PRIOR APPLICATION NUMBER: 60/179,065
;; PRIOR FILING DATE: 2000-01-31
;; PRIOR APPLICATION NUMBER: 60/180,628
;; PRIOR FILING DATE: 2000-02-04
;; PRIOR APPLICATION NUMBER: 60/209,467
;; PRIOR FILING DATE: 2000-06-07
;; NUMBER OF SEQ ID NOS: 167
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO 145
;; LENGTH: 20272
;; TYPE: DNA
;; ORGANISM: Homo sapiens
US-09-908-711-145

Query Match 1.2%; Score 18; DB 10; Length 20272;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1153 TCAGAGCTGCTGAAATG 1170

Db 6642 TCAGAGCTGTCTGAAATG 6625
|||||
RESULT 15
US-09-908-711-143/c
; Sequence 143, Application US/09908711
; Patent No. US20020045230A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA128
; CURRENT APPLICATION NUMBER: US/09/908,711
; CURRENT FILING DATE: 2001-07-20
; PRIOR APPLICATION NUMBER: US01/01360
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,867
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01344
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,892
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01345
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,888
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01329
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,905
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01354
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,891
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01339
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,869
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01340
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,874
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01334
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,898
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01320
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,853
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01349
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,902
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01239
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,870
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01348
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,882
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01347
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,896
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01307
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,864
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01341
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,856

; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01336
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,868
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01312
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: 60/209,467
; PRIOR FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 167
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 143
; LENGTH: 25619
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-908-711-143
Query Match 1.2%; Score 18; DB 10; Length 25619;
Best Local Similarity 100.0%; Pred.No. 13;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1153 TCAGAGCTGTCTGAAATG 1170
|||||
Db 1048 TCAGAGCTGTCTGAAATG 1031

Search completed: April 8, 2003, 23:20:39
Job time : 285 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 8, 2003, 20:42:04 ; Search time 2260 Seconds
(without alignments)
10462.571 Million cell updates/sec

Title: US-09-595-947C-1

Perfect score: 1460

Sequence: 1 gcaggtacgagagagcag.....agagtgaacctaatccagtgt 1460

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 16154066 seqs, 8097743376 residues

Word size : 15

Total number of hits satisfying chosen parameters: 29042

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

```
EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_estc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pin:*
21: em_gss_vit:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	92	6.3	593	17	AZ296526
2	60	4.1	1540	11	AK008017
3	23	1.6	600	12	BG808248
4	23	1.6	804	9	AL540071
5	23	1.6	814	14	BQ178789
6	23	1.6	1269	12	BG854922

7	1.4	319	12	BE936551
8	1.4	464	17	AQ753599
9	1.4	536	17	AZ508899
10	1.4	718	10	BE263765
11	1.4	829	13	BI910102
12	1.4	913	12	BF204175
13	1.4	953	9	AU067624
14	1.4	962	14	BQ686909
15	1.4	1022	13	BM563662
16	1.4	1039	12	BE780690
17	1.4	1042	12	BG419220
18	1.4	241	9	AA113743
19	1.4	351	14	T70947
20	1.4	397	13	BI445873
21	1.4	444	10	AV944801
22	1.4	476	13	BM253695
23	1.4	550	10	AV939670
24	1.4	571	17	CNS01YSD
25	1.4	971	17	CNS02YD5
26	1.4	1075	14	BM810917
27	1.3	201	10	BB071548
28	1.3	234	10	BB269747
29	1.3	260	10	BE068280
30	1.3	301	10	BB305111
31	1.3	366	10	AW426722
32	1.3	370	10	AW481542
33	1.3	468	17	AZ286451
34	1.3	496	12	BF191528
35	1.3	507	12	BF408307
36	1.3	529	17	AZ830930
37	1.3	533	17	AZ985482
38	1.3	541	9	AU057600
39	1.3	552	13	BM030263
40	1.3	555	10	BE600721
41	1.3	564	13	BM088009
42	1.3	570	17	AQ594075
43	1.3	577	10	AW672224
44	1.3	579	10	AV597322
45	1.3	585	17	AQ631359

ALIGNMENTS

RESULT 1
AZ296526/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

AZ296526
RPCI-23-160G18-TV RPCI-23 Mus musculus genomic clone RPCI-23-160G18
DNA sequence.
593 bp
DNA
linear
GSS 27-JUL-2000

AZ296526
GI:9538311
GSS.
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 593)
Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret
,B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Kroi,M., de Jong,P.
and Fraser,C.M.
Mouse BAC End Sequences from Library RPCI-23
Unpublished (1999)
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from


```

/strain="C57BL/6J"
/db_xref="FANTOM_DB:2010001M19"
/db_xref="MGI:1907403"
/db_xref="taxon:10090"
/clone="2010001M19"
/sex="male"
/tissue_type="small intestine"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
1..1540
/gene="Atoh5"
241..885
/gene="Atoh5"
/note="atonal homolog 5 (Drosophila)"
data source:MGI, source key:MGI:893591, evidence:ISS
putative"
/codon_start=1
/protein_id="BAB25411.1"
/db_xref="GI:12841942"
/db_xref="MGI:893591"
/translation="MAPHLDALTIQVSPETQPPGASDHEVLSSNTPPSHTLIPR
DCSEAEVDCGTSRKLRRAGRRPKSELALSKRRSRKKANDRRNRNHNLSA
LDALRGVLPFTDDAKLVETRFANHYIWAQTQTLRIADHSXGPPVPCGELGS
PGGSGNGDWGSIYSPVSOAGNLSPTRASLEFPGLQVPSPSYLLPGLVFSDFL"
BASE COUNT      306 a 485 c 397 g 352 t
ORIGIN
Query Match      4.1%; Score 60; DB 11; Length 1540;
Best Local Similarity 100.0%; Pred. No. 4.8e-20;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 762 GCGCTGGATGCGTGGCGGTGTCCTGCCACCTTCCCGATGACGCCAACCTTACAAG 821
|||||
Db 544 GCGCTGGATGCGTGGCGGTGTCCTGCCACCTTCCCGATGACGCCAACCTTACAAG 603

RESULT 3
BG808248
LOCUS      600 bp mRNA linear EST 20-DEC-2001
DEFINITION      2083-52 Mouse E14.5 retina lambda ZAP II Library Mus musculus cDNA,
mRNA sequence.
ACCESSION      BG808248
VERSION      BG808248.1 GI:17955225
KEYWORDS      EST.
SOURCE      house mouse.
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 600)
REFERENCE      Mu, X., Zhao, S., Pershad, R., Hsieh, T.-F., Scarpa, A., Wang, S.W.,
White, R.A., Beremand, P.D., Thomas, T.L., Gan, L. and Klein, W.H.
Gene expression in the developing mouse retina by EST sequencing
and microarray analysis
JOURNAL      Nucleic Acids Res. 29 (24), 4983-4993 (2001)
MEDLINE      21671825
COMMENT      Contact: Klein WH
Department of Biochemistry and Molecular Biology
University of Texas M.D. Anderson Cancer Center
Box 117, 1515 Holcombe Blvd., Houston, TX 77030, USA
Tel: 713 792 3646
Fax: 713 790 0329.

FEATURES
source
1..600
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone_lib="Mouse E14.5 retina lambda ZAP II Library"
/tissue_type="neural retina"
/dev_stage="embryonic day 14.5 post-fertilization"
/notes="Vector: pAMP10 (Gibco); Cloned unidirectionally.
Primer: Oligo dt. RNA Isolation: cytoplasmic RNA preps
(Mannatis); Cloning Technique: CUA Cloning (cloneAmp,
Life Technologies); Average insert size: 1.8 Kb;
Insertion site: TACGTCCTCACTGAATCTGAGTG---. Other

BASE COUNT      154 a 270 c 278 g 99 t
ORIGIN
Query Match      1.6%; Score 23; DB 9; Length 804;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 831 CTGCGCTTCGCCACCACTACAT 853
|||||
Db 770 CTGCGCTTCGCCACCACTACAT 792

RESULT 5
BQ178789
LOCUS      814 bp mRNA linear EST 30-APR-2002
DEFINITION      UI-M-EVO-bwt-k-07-0-UI.rl NIH-BMAP_EVO Mus musculus cDNA clone
IMAGE:5701758 5', mRNA sequence.
ACCESSION      BQ178789
VERSION      BQ178789.1 GI:20354281

```

Information regarding entire library may be found at
http://pga.swmed.edu/Data/Libraries/microarray_cdna_libraries.htm.

```

BASE COUNT      94 a 238 c 161 g 107 t
ORIGIN
Query Match      1.6%; Score 23; DB 12; Length 600;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 732 GAGCGCAACCGCATGCACAACT 754
|||||
Db 4 GAGCGCAACCGCATGCACAACT 26

RESULT 4
AL540071
LOCUS      804 bp mRNA linear EST 16-FEB-2001
DEFINITION      AL540071 LTI_FLO13_FBrn1 Homo sapiens cDNA clone CS0DF035YD19 5
prime, mRNA sequence.
ACCESSION      AL540071
VERSION      AL540071.1 GI:12869886
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 804)
REFERENCE      Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
JOURNAL      Unpublished (2001)
COMMENT      Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
source
1..804
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DF035YD19"
/clone_lib="LTI_FLO13_FBrn1"
/dev_stage="pooled tissue from post conception fetuses (20
week, 24 week and 26 week)"
/lab_host="DH10B"
/notes="Organ: Fetal brain; Vector: pCMVSPORT 6; 1st strand
cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-stranded cDNA was digested with Not I
and cloned into the Not I and Eco RV sites of the
pCMVSPORT 6 vector. Library was constructed by Life
Technologies. Contact : Feng Liang Life Technologies, a
division of Invitrogen 9800 Medical Center Drive Rockville
Maryland 20850, USA Fax : (1) 301 610 8371 Email :
filiang@lifetech.com URL :
http://fulllength.invitrogen.com"

BASE COUNT      154 a 270 c 278 g 99 t
ORIGIN
Query Match      1.6%; Score 23; DB 9; Length 804;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 831 CTGCGCTTCGCCACCACTACAT 853
|||||
Db 770 CTGCGCTTCGCCACCACTACAT 792

RESULT 5
BQ178789
LOCUS      814 bp mRNA linear EST 30-APR-2002
DEFINITION      UI-M-EVO-bwt-k-07-0-UI.rl NIH-BMAP_EVO Mus musculus cDNA clone
IMAGE:5701758 5', mRNA sequence.
ACCESSION      BQ178789
VERSION      BQ178789.1 GI:20354281

```


This sequence was derived from the FAPESP/LICR Human Cancer Genome project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=st2=RC5-NT0053-310
 800-024-E01st3-2000-08-31&t4=1)
 Seq primer: puc 18 forward
 High quality sequence stop: 319.

FEATURES

Location/Qualifiers
 1..319
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="NT0053"
 /dev_stage="Adult"
 /note="Organ: nervous tumor; Vector: puc18; Site: 1; SmaI;
 Site: 2; SmaI; A mini-library was made by cloning products
 derived from ORESTES PCR (U.S. Letters Patent application
 No. 196,716 - Ludwig Institute for Cancer Research)
 profiles into the pUC 18 vector. Reverse transcription of
 tissue mRNA and cDNA amplification were performed under
 low stringency conditions."

BASE COUNT

53 a 87 c 106 g 73 t

ORIGIN

Query Match 1.4%; Score 21; DB 12; Length 319;
 Best Local Similarity 100.0%; Pred. No. 15;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1112 AACAGGCCGTGGCGGTGGGC 1132

|||||

Db 280 AACAGGCCGTGGCGGTGGGC 300

RESULT 8

AQ753599

LOCUS

HS 2117_A2_H03_MR CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate-2117 Col-6 Row-O, DNA sequence.

ACCESSION

AQ753599

VERSION

AQ753599.1

KEYWORDS

GSS.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 464)

Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,

Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and

Hood,L.

Sequence-tagged connectors: A sequence approach to mapping and

scanning the human genome

Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)

99380589

Contact: Mahairas GG, Wallace JC, Hood L

High Throughput Sequencing Center

University of Washington

401 Queen Anne Avenue North, Seattle, WA 98109, USA

Tel: (206) 616-3618

Fax: (206) 616-3887

Email: jwallace@u.washington.edu

Clones may be purchased from Research Genetics (info@resgen.com).

BAC end Web Server: http://www.htsc.washington.edu

Plate: 2117 row: 0 column: 6

Seq primer: M13 Reverse

Class: BAC ends

High quality sequence stop: 464.

Location/Qualifiers

1..464

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="Plate-2117 Col-6 Row-O"

/clone_lib="CIT Approved Human Genomic Sperm Library D"

/sex="male"

/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in

E-Coli DH10B"

FEATURES

source

BASE COUNT 125 a 114 c 101 g 123 t 1 others

ORIGIN

Query Match 1.4%; Score 21; DB 17; Length 464;

Best Local Similarity 100.0%; Pred. No. 16;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 492 CAAAGTGTCCCAAGAGACCCAG 512

|||||

Db 106 CAAAGTGTCCCAAGAGACCCAG 126

RESULT 9

AZ506899

LOCUS

IM0348G13F Mouse 536 bp DNA linear GSS 05-OCT-2000 clone UUGC1M0348G13 F, DNA sequence.

ACCESSION

AZ506899

VERSION

AZ506899.1

KEYWORDS

GSS.

SOURCE

house mouse.

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 536)

Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,

Islam,H., Longacre,S., Mahmood,M., Meenen,E., Pedersen,T., Reilly

,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.

and Wright,D., Weiss,R.

Mouse whole genome scaffolding with paired end reads from 10kb

plasmid inserts

Unpublished (2000)

Contact: Robert B. Weiss

University of Utah Genome Center

University of Utah

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT

84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: rdunn@genetics.utah.edu

Insert Length: 10000

Plate: 0348 row: G column: 13

Seq primer: CGTGTAAACGACGGCCAGT

Class: plasmid ends

High quality sequence stop: 536.

Location/Qualifiers

1..536

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="UUGC1M0348G13"

/clone_lib="Mouse 10kb plasmid UUGC1M library"

/sex="Male"

/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"

/note="Vector: PWD42nv; Purified genomic DNA from M.

musculus C57BL/6J (male) was obtained from the Jackson

Laboratory Mouse DNA Resource

(http://www.jax.org/resources/documents/dnares/). The DNA

was hydrodynamically sheared by repeated passage through a

0.005 inch orifice at constant velocity. The sheared DNA

was blunt end-repaired with T4 DNA polymerase and T4

polynucleotide kinase. Adaptor oligonucleotides were

ligated to the blunt ends in high molar excess. The

adapted DNA was purified and size-selected for a 9.5 to

10.5 kb range using preparative agarose gel

electrophoresis. Vector DNA was prepared from a derivative

of pWD42 (g114732114|gb|AF129072.1), a copy-number

inducible derivative of plasmid RL. The vector was ligated

with adaptors complementary to the insert adaptors and

purified. The sheared, adapted mouse DNA was annealed to

adapted vector DNA, and transformed into

chemically-competent E. coli XL10-Gold (Stratagene) cells

and selected for ampicillin resistance.

```

BASE COUNT      156 a      119 c      81 g      180 t
ORIGIN

Query Match      1.4%; Score 21; DB 17; Length 536;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 916 TSCCCTGTGGGAGCTGGGAA 936
|||||
Db 406 TGCCCTGTGGGAGCTGGGAA 426

RESULT 10
BE263765
LOCUS 601194122F1 NIH_MGC_7 Homo sapiens cDNA clone IMAGE:5337940 5',
DEFINITION mRNA sequence.
ACCESSION BE263765
VERSION BE263765.1 GI:9137311
KEYWORDS EST.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 718)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Plate: LICM222 row: 1 column: 05
High quality sequence stop: 649.
FEATURES
source
Location/Qualifiers
1..718
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3537940"
/tissue_type="small cell carcinoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: lung; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
```

```

BASE COUNT      115 a      195 c      260 g      148 t
ORIGIN

Query Match      1.4%; Score 21; DB 10; Length 718;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1112 AACAGGCCCTGGCGGTGGC 1132
|||||
Db 536 AACAGGCCCTGGCGGTGGC 556

RESULT 11
BI910102
LOCUS B1910102 829 bp mRNA linear EST 16-OCT-2001
DEFINITION 603067946F1 NIH_MGC_118 Homo sapiens cDNA clone IMAGE:5217154 5',
mRNA sequence.
ACCESSION BI910102
VERSION BI910102.1 GI:16173443
KEYWORDS EST.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 913)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LICM996 row: a column: 07
High quality sequence stop: 714.
FEATURES
source
Location/Qualifiers
1..913
/organism="Homo sapiens"

REFERENCE 1 (bases 1 to 829)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM11545 row: k column: 11
High quality sequence stop: 828.
FEATURES
source
Location/Qualifiers
1..829
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5217154"
/tissue_type="leukocyte"
/lab_host="DH10B"
/notes="Vector: PCMV-SPORT6; Site_1: NotI; Site_2: EcoRV
(destroyed); RNA source leukocytes from anonymous pool of
non-activated adult donors. Library is oligo-dT primed
and directionally cloned (EcoRV site is destroyed upon
cloning). Average insert size 1.7 kb, insert size range
1.2-3.3 kb. Library is normalized and enriched for
full-length clones and was constructed by C. Gruber
(invitrogen). Research Genetics tracking code 027. Note:
this is a NIH-MGC Library."
BASE COUNT      146 a      224 c      293 g      166 t
ORIGIN

Query Match      1.4%; Score 21; DB 13; Length 829;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1112 AACAGGCCCTGGCGGTGGC 1132
|||||
Db 550 AACAGGCCCTGGCGGTGGC 570

RESULT 12
BF204175
LOCUS BF204175 913 bp mRNA linear EST 06-NOV-2000
DEFINITION 601867625F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:4110222 5',
mRNA sequence.
ACCESSION BF204175
VERSION BF204175.1 GI:11097761
KEYWORDS EST.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 913)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LICM996 row: a column: 07
High quality sequence stop: 714.
FEATURES
source
Location/Qualifiers
1..913
/organism="Homo sapiens"
```


KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1022)
AUTHORS NIH-MGC <http://mgs.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LCM1916 row: k column: 15
High quality sequence stop: 654.
Location/Qualifiers
1..1022
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5441798"
/clone_lib="NIH_MGC_98"
/tissue_type="astrocytoma grade IV, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Library constructed by Ling Hong
in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."

BASE COUNT 177 a 324 c 319 g 202 t
ORIGIN
Query Match 1.4%; Score 21; DB 13; Length 1022;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1112 AACAGGCCCTGGCGGTGGC 1132
|||||
Db 573 AACAGGCCCTGGCGGTGGC 593.

Search completed: April 8, 2003, 23:14:59
Job time : 2274 secs

GenCore version 5.1.4_p5.4578
Copyright(c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 8, 2003, 21:20:34 ; Search time 2813 Seconds
(without alignments)
2214.007 Million cell updates/sec

Title: US-09-595-947C-10

Perfect score: 1127

Sequence: 1 MTPQSGATVQVTRTERS.....LGATSSACLSPGSLAFSDFL 214

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xilh
-Q/cgn2.1/USPTO.spool/US09595947/runat_07042003_160404_20061/app_query.fasta_1.391
-DB=GenEmbl -OPMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTPMT=pto -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09595947 -CGN_1_1_2496_@runat_07042003_160404_20061 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MAP -LARGQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.*

1: gb_ba.*
2: gb_htg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*

29: em_vi.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htg_mus.*
34: em_htg_pln.*
35: em_htg_rod.*
36: em_htg_mam.*
37: em_htg_vrt.*
38: em_sy.*
39: em_htgo_hum.*
40: em_htgo_mus.*
41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1127	100.0	1330	9	HSAL33776
2	1105	98.0	5340	9	AF234829
3	1099	97.5	165110	9	AL450311
4	1089	96.6	173341	2	AC021954
5	849	75.3	1491	6	A91167
6	849	75.3	1491	10	RNRELAXT
7	819	72.7	861	10	MMU76208
8	819	72.7	1861	10	AF364300
9	819	72.7	5567	10	MMATH4B
10	819	72.7	138070	2	AC109783
11	819	72.7	185806	2	AC127417
12	523	46.4	170896	2	AC011010
13	388.5	34.5	790	5	GGA012659
14	388.5	34.5	1880	5	AF303000
15	387.5	34.4	105572	2	AC112007
16	385.5	34.2	1074	5	GGA012660
17	382	33.9	1299	5	AF123883
18	381.5	33.9	770	5	AF123884
19	380.5	33.8	1341	5	AF109014
20	378.5	33.6	1268	6	AR023709
21	378.5	33.6	1675	9	BC008687
22	378.5	33.6	1717	9	BC028226
23	378.5	33.6	79970	2	AC011428
24	378.5	33.6	134506	9	AC005738
25	376.5	33.4	1268	9	HSU63842
26	375.5	33.3	1385	10	MMU76207
27	375.5	33.3	10393	10	AF303001
28	370.5	32.9	1412	10	MMATH4A
29	370.5	32.9	151044	2	AC011432
30	370	32.8	1527	10	RN067777
31	370	32.8	16677	2	AC127817
32	368.5	32.7	123855	2	AC102600
33	365	32.4	932	10	MMATH4C
34	365	32.4	1315	10	MMU63841
35	365	32.4	1333	6	AR023715
36	365	32.4	174688	2	AC124395
37	359.5	31.9	735	10	MMU67776
38	359.5	31.9	71538	2	AC118243
39	356.5	31.6	1719	5	AB065284
40	347.5	30.8	91531	2	AC079846_3
41	347.5	30.8	179697	9	AC023886
42	346	30.7	6123	9	AF303002
43	329.5	29.2	1176	3	AF271788
44	327	29.0	750	5	DRU94588
45	327	29.0	1098	5	AF024535

ALIGNMENTS

RESULT 1

```

HSA133776
LOCUS HSA133776 1330 bp DNA linear PRI 19-JUN-1999
DEFINITION Homo sapiens gene for neurogenin 3.
ACCESSION AJ133776
VERSION AJ133776.1 GI:5123782
KEYWORDS bHLH transcription factor; neurogenesis; neurogenin 3; ngn3 gene.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1330)
AUTHORS Ravassard,P., Icard-Liepkains,C., Wiard,L., Julien,J.P. and Mallet,J.
TITLE The human neurogenin 3 homolog maps to chromosome 10q21.3 and its expression pattern is identical to that of its murine counterparts
JOURNAL Unpublished
AUTHORS Ravassard,P.
TITLE Direct Submission
JOURNAL Submitted (16-MAR-1999) Ravassard P., Lgn, CNRS UMRC 9923, Hopital de la Pitie Salpetriere, Bat. CERVI, 83 Bd. de l'Hopital, 75013 PARIS, FRANCE
FEATURES
source Location/Qualifiers
1..1330
/organism="Homo sapiens"
/db_xref="taxon:9606"
gene /gene="ngn3"
mRNA /gene="ngn3"
join(<1..157,321..>1330)
exon 1..157
/gene="ngn3"
intron 158..320
/gene="ngn3"
number=1
exon 321..1330
/gene="ngn3"
number=2
CDS 322..966
/gene="ngn3"
/function="early neurogenesis"
/codon_start=1
/product="neurogenin 3"
/protein_id="CA45384.1"
/db_xref="GI:5123783"
/translation="MTPOPSGAPTVOVTRERSPPRASEDEVCTSPAPSPPTPG
NCAAEBSGGCGAPKRLARCGRSRPSKSELSKQRRSRKANDRENRHDLNSA
LDALGVLPFPDDAKLKIETLFAHNYIWAOTLRIADHLSLVALEPPHAPGCELG
SPGPGPGWGLSYSPVQAGSLSPAAISLEERPLGLIGATSSACLSPGSLAFSDFL"
BASE COUNT 230 a 459 c 413 g 228 t
ORIGIN
Alignment Scores:
Pred. No.: 2,26e-46 Length: 1330
Score: 1127.00 Matches: 214
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0
US-09-595-947C-10 (1-214) x HSA133776 (1-1330)
QY 1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
Db 322 ATGACCCCTCAACCTTCGGGTGGCCCACTGTCCAAAGTGACCGCTCGCCCGCCGAGCGGTCC 381
QY 21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40
Db 382 TTCGCCAGACCTCGGAACAGAGTACCTGCCACAGCTCGCCCGCCGAGCGGTCC 441
QY 41 ArgThrProGlyAsnCysAlaGluAlaGluGluGlyCysArgGlyAlaProArgLys 60

```


LDALRGVLPDPDAKLTETLRFHNYIWTQTLRIADHSLYALEPAPHCGLG
SPGSGPDGSLYSPVSOAGSLSPAASLERPGLIGTSSACLSPGSLAFSDFL"
BASE COUNT 1215 a 1508 c 1514 g 1111 t
ORIGIN

Alignment Scores:
Pred. No.: 8.92e-45 Length: 5340
Score: 1105.00 Matches: 211
Percent Similarity: 99.07% Conservative: 1
Best Local Similarity: 98.60% Mismatches: 2
Query Match: 98.05% Indels: 0
DB: 9 Gaps: 0

US-09-595-947c-10 (1-214) x AF234829 (1-5340)

Qy 1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
Db 3022 ATGACGGCTCAACCTCGGGTGCCTCCACTGTCCAAAGTGACCCGTGAGACGGAGGTCC 3081
Qy 21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40
Db 3082 TTCCCCAGAGCCTCGGAAGCAAGTAGTACCTGCCCCACGTCGCCGCCGCCGCCCACT 3141
Qy 41 ArgThrProGlyAsnCysAlaGluAlaGluGluGlyCysArgGlyAlaProArgLys 60
Db 3142 CGCACAGGGGGAAGTCCGACAGAGGGGAGAGGAGGCTCCGAGGGGCCCCAGGAG 3201
Qy 61 LeuArgAlaArgArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80
Db 3202 CTCGGGCACGCGCGGGGACGACGCGGCTTAAGAGCGAGTGGCACTGAGCAAGCAG 3261
Qy 81 ArgArgSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn 100
Db 3262 CGACGGAGTCGGCGAAGAGCCCAACGACCGGCGGCAATCGAATCCACAACCTCAAC 3321
Qy 101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThr 120
Db 3322 TCGGCACTGGACGCGCTCGCGGTGTCTGCCCACCTTCCAGACGAGCGGAGCTCAC 3381
Qy 121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArg 140
Db 3382 AAGATCGAGAGCGTGCCTTCGCCACAACTACATCTGGGCGCTGACTCAAAACGCTGCG 3441
Qy 141 IleAlaAspHisSerLeuThrAlaLeuGluProAlaProHisCysGlyGluLeuGly 160
Db 3442 ATAGCGGACCAAGCTTGTACGCGCTGGAGCGCGCGCGGCTGCGGGGAGCTGGGC 3501
Qy 161 SerProGlyGlyProProGlyAspTrpGlySerLeuThrSerProValSerGlnAlaGly 180
Db 3502 AGCCAGGCGGTTCCTCCCGGGAGCTGGGGTCCCTCTACTCCAGTCTCCAGGCTGGC 3561
Qy 181 SerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeuLeuGlyAlaThrSerSer 200
Db 3562 AGCCTGAGTCCCGCGCTCGCTGGAGAGGAGCGGCGGCTGCTGGGGGCCACCTCTTCC 3621
Qy 201 AlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214
Db 3622 GCCTGCTTGAGCCGAGGAGCTGCTGCTTCTCAGATTTCG 3663

RESULT 3
AL450311/c 165110 bp DNA linear PRI 12-JUL-2001
LOCUS Human DNA sequence from clone RP11-343J3 on chromosome 10, complete
DEFINITION sequence.
ACCESSION AL450311
VERSION AL450311.11 GI:14626972
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 165110)
AUTHORS Howden, P.

TITLE JOURNAL

Direct Submission
Submitted (12-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
requests: clonerequest@sanger.ac.uk

COMMENT

On Jul 8, 2001 this sequence version replaced gi:14575291.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em, EMBL; SW,
SWISSPROT; Tr, TREMBL; Wp, WORMPEP; information on the WORMPEP
database can be found at

http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 10, constructed by the Sanger Centre Chromosome 10
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr10

RP11-343J3 is from the library RP11-11.2 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm

VECTOR: pBAC3.6

This sequence is the entire insert of clone RP11-343J3 The true
left end of clone RP11-242G20 is at 139955 in this sequence. The
true right end of clone RP11-404C6 is at 6588 in this sequence.

FEATURES

source

1. 165110
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-343J3"
/clone_lib="RP11-11.2"
7. 147

repeat_region /note="THE1C repeat: matches 2. 142 of consensus"
151. 293
repeat_region /note="MIR repeat: matches 2. 148 of consensus"
1136. 1187
repeat_region /note="26 copies 2 mer gt 98% conserved"
2960. 3272
repeat_region /note="AluSq repeat: matches 1. 310 of consensus"
4144. 4270
repeat_region /note="AluSx repeat: matches 1. 134 of consensus"
4319. 4494
repeat_region /note="AluSx repeat: matches 118. 293 of consensus"
5662. 6438
repeat_region /note="L1MEC repeat: matches 272. 1095 of consensus"
6916. 8080
repeat_region /note="L1MEC repeat: matches 1168. 2367 of consensus"
9934. 10087
repeat_region /note="MIR repeat: matches 25. 184 of consensus"
10353. 10481
repeat_region /note="MIR repeat: matches 65. 194 of consensus"
11077. 11372
repeat_region /note="AluSx repeat: matches 1. 294 of consensus"
11600. 11639
repeat_region /note="20 copies 2 mer tc 95% conserved"
12479. 12608
repeat_region /note="65 copies 2 mer at 90% conserved"
13433. 13876
repeat_region /note="MLTIC repeat: matches 1. 466 of consensus"
14372. 14576
repeat_region /note="AluSq repeat: matches 1. 301 of consensus"
14577. 14633
repeat_region /note="AluSq repeat: matches 243. 299 of consensus"

```

repeat_region 14843..14943
/note="MER86 repeat: matches 9. .109 of consensus"
repeat_region 1504..415613
/note="MIR repeat: matches 3. .213 of consensus"
repeat_region 16626..16705
/note="MIR repeat: matches 60. .140 of consensus"
repeat_region 16928..16993
/note="33 copies 2 mer gg 66% conserved"
repeat_region 17476..17569
/note="MER81 repeat: matches 2. .114 of consensus"
repeat_region 17719..18069
/note="LMC4 repeat: matches 7617. .7977 of consensus"
repeat_region 18117..18345
/note="MIR repeat: matches 7. .262 of consensus"
repeat_region 19169..19400
/note="MIR repeat: matches 5. .239 of consensus"
repeat_region 19436..19627
/note="LMC5 repeat: matches 7728. .7917 of consensus"
repeat_region 19642..19716
/note="MIR repeat: matches 48. .131 of consensus"
repeat_region 19969..20010
/note="21 copies 2 mer tg 100% conserved"
repeat_region 20623..20704
/note="41 copies 2 mer gt 85% conserved"
repeat_region 20738..20773
/note="U2 repeat: matches 1. .36 of consensus"
repeat_region 20802..20863
/note="MLT1J repeat: matches 1. .62 of consensus"
repeat_region 20889..21263
/note="MLT1F repeat: matches 188. .541 of consensus"
repeat_region 21463..21618
/note="MIR repeat: matches 46. .192 of consensus"
repeat_region 22019..22326
/note="AluJb repeat: matches 1. .306 of consensus"
repeat_region 22381..22564
/note="L2 repeat: matches 2453. .2629 of consensus"
repeat_region 22896..23174
/note="MLT1J repeat: matches 117. .413 of consensus"
repeat_region 23215..23346
/note="MIR repeat: matches 48. .188 of consensus"
repeat_region 23388..23532
/note="L2 repeat: matches 2097. .2230 of consensus"
repeat_region 23533..23837
/note="AluJb repeat: matches 1. .303 of consensus"
repeat_region 23838..24137
/note="L2 repeat: matches 1754. .2097 of consensus"
repeat_region 24291..24581
/note="AluX repeat: matches 1. .300 of consensus"
repeat_region 24653..24850
/note="MIR repeat: matches 1. .200 of consensus"
misc_feature 26189..26685
/note="Cpg island"
/evidence=not_experimental
misc_feature 29453..30918
/note="Cpg island"
/evidence=not_experimental
repeat_region 31084..31156
/note="MIR repeat: matches 65. .138 of consensus"
repeat_region 31618..31876
/note="AluJo repeat: matches 29. .275 of consensus"
repeat_region 32767..32830
/note="MIR repeat: matches 76. .139 of consensus"
repeat_region 33050..33178
/note="43 copies 3 mer tcc 72% conserved"
misc_feature 35112..36201
/note="Cpg island"
/evidence=not_experimental
repeat_region 36551..36604
/note="27 copies 2 mer ac 94% conserved"
repeat_region 37585..38254
/note="LIPAL5 repeat: matches 5480. .6157 of consensus"
repeat_region 38265..38561
/note="AluX repeat: matches 1. .300 of consensus"
repeat_region 39214..39423
/note="MIR repeat: matches 3. .219 of consensus"
repeat_region 39489..39801
/note="AluX repeat: matches 1. .292 of consensus"
repeat_region 40050..40189
/note="MIR repeat: matches 1. .144 of consensus"
repeat_region 41057..41290
/note="MIR repeat: matches 8. .240 of consensus"
repeat_region 41296..41460
/note="L2 repeat: matches 2569. .2730 of consensus"
repeat_region 41506..41944
/note="L2 repeat: matches 1916. .2416 of consensus"
repeat_region 42388..42698
/note="L2 repeat: matches 1448. .1779 of consensus"
repeat_region 44193..44579
/note="THE1C repeat: matches 1. .371 of consensus"
repeat_region 44600..44956
/note="LTR16A repeat: matches 90. .445 of consensus"
repeat_region 45240..45300
/note="MER58A repeat: matches 37. .97 of consensus"
repeat_region 45798..45909
/note="LTR41 repeat: matches 90. .192 of consensus"
repeat_region 46826..46871
/note="23 copies 2 mer gt 97% conserved"
repeat_region 50584..50969
/note="AluSq repeat: matches 1. .287 of consensus"
repeat_region 50980..51291
/note="AluX repeat: matches 1. .312 of consensus"
repeat_region 52222..52519
/note="AluX repeat: matches 1. .300 of consensus"
repeat_region 54065..54260
/note="LM4 repeat: matches 3865. .4055 of consensus"
repeat_region 54261..54432
/note="FAM repeat: matches 2. .167 of consensus"
repeat_region 54433..54629
/note="LM4 repeat: matches 3652. .3865 of consensus"
repeat_region 54648..54862
/note="LTR41 repeat: matches 11. .217 of consensus"
repeat_region 54863..55236
/note="MLT1A1 repeat: matches 1. .365 of consensus"
repeat_region 55237..55700
/note="LTR41 repeat: matches 217. .716 of consensus"

```

Alignment Scores:

Pred. No.:	3.72e-43	Length:	165110
Score:	1099.00	Matches:	210
Percent Similarity:	98.60%	Conservative:	1
Best Local Similarity:	98.13%	Mismatches:	3
Query Match:	97.52%	Indels:	0
DB:	9	Gaps:	0

US-09-595-947C-10 (1-214) x AL450311 (1-165110)

Qy 1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
 |||||
 Db 30325 ATGACGGCTCACCCTCGGGTGGCCCTGTCCAAAGTACCGCTGACGAGGCGGGTCC 30266
 |||||

Qy 21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40
 |||||
 Db 30265 TTCCCCAGAGCTCGGAAGACGAAGTACCTGCCACGCTCCGCCGCCGCCGCCACT 30206
 |||||

Qy 41 ArgThrProGlyAsnCysAlaGluAlaGluGlyCysArgGlyAlaProArgLys 60
 |||||
 Db 30205 CGCACACGGGGGAACCTCGCAGAGCGGAGGAGGCTCCGAGGGGCCCGCCGAGGAAG 30146
 |||||

Qy 61 LeuArgAlaArgArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80
 |||||
 Db 30145 CTCGGGCACGGCGCGGGGAGCGGACGCGCGGCTTAAGAGCGAGTGGCACTGAGCAACGAG 30086
 |||||

Qy 81 ArgArgSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn 100
 |||||
 Db 30085 CGACGGAGTCGGCGAAAGAGGCCAACGACCGCGAGCGCAATCAATGACAACTCAAC 30026
 |||||

QY 101 SerAlaLeuAspAlaLeuArgglyValLeuProThrPhleProAspAspAlaLysLeuThr 120
 |||||
 Db 30025 TGGGACTGGAGCGCTGGCGTGTCTCCACACCTCCAGACGACGCGAAGCTCAC 29966
 |||||
 QY 121 LysIleGluThrLeuArgPheAlaHisAanTyrIleTrrPAlaLeuThrGlnThrLeuArg 140
 |||||
 Db 29965 AAGATCGAGACGCTGGCTTCCGCCACAACTACATCTGGGCGCTGACTCAACGCGTGGC 29906
 |||||
 QY 141 IleAlaAspHisSerLeuTyrAlaLeuGluProProAlaProHisCysGlyGluLeuGly 160
 |||||
 Db 29905 ATAGCGGACACAGCTGTACCGCTGGAGCGCGCGCGCCGACCTGCGGGAGCTGGGC 29846
 |||||
 QY 161 SerProGlyGlyProProGlyAspTrpGlySerLeuTyrSerProValSerGlnAlaGly 180
 |||||
 Db 29845 AGCCAGGCGGTTCGCCCGGAGCTGGGGTCCCTCTACTCCCACTCTCCAGGCTGGC 29786
 |||||
 QY 181 SerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeuLeuGlyAlaThrSerSer 200
 |||||
 Db 29785 AGCCTGAGTCCGCGCGCTGCTGGAGGAGCGACCGCGGCTGCTGGGGCCACCTTTCC 29726
 |||||
 QY 201 AlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214
 |||||
 Db 29725 GCCTGTTGAGCCGAGGCTGCTGGCTTCTCAGATTCTG 29684
 |||||

RESULT 4
 AC021954/c
 LOCUS Homo sapiens chromosome 10 clone RP11-57E12 map 10, WORKING DRAFT
 DEFINITION AC021954
 ACCESSION AC021954
 VERSION AC021954.3 GI:7417809
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 173341)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
 Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
 Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
 Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
 DeRellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
 Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
 Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
 Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
 Landers,T., Lehoczy,J., Levine,R., Liu,C., Liu,G., Locke,K.,
 Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
 McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
 Norman,C.H., O'Connor,T., O'Donnell,P., Oliver,T.M., Peterson,K.,
 Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
 Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
 Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
 Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
 Zimmer,A. and Zody,M.
 Direct Submission
 Submitted (22-JAN-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 3 (bases 1 to 173341)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
 Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
 Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G.,
 Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
 Collymore,A., Cooke,P., DeRellano,K., Dewar,K., Diaz,J.S.,
 Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
 Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
 Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
 Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
 Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczy,J.,
 Levine,R., Liu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,

TITLE
 JOURNAL
 COMMENT

McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
 Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mieng,V., Morrow,J.,
 Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
 O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
 Pisani,C., Pollara,V., Raymond,C., Riley,R., Roqov,P., Rothman,D.,
 Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
 Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
 Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
 Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
 Young,G., Zainoun,J., Zimmer,A. and Zody,M.
 Direct Submission
 Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Apr 5, 2000 this sequence version replaced gi:6984451.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L5931
 Center clone name: 57_E12
 ----- Summary Statistics
 Sequencing vector: M13; M7815; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 161190 bases at least Q40
 Consensus quality: 166837 bases at least Q30
 Consensus quality: 168995 bases at least Q20
 Insert size: 176000; agarose-1p
 Insert size: 171041; sum-of-contigs
 Quality coverage: 3.7 in Q20 bases; agarose-1p
 Quality coverage: 3.8 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 24 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 1574: contig of 1574 bp in length
 * 1575 1674: gap of 100 bp
 * 1675 3043: contig of 1369 bp in length
 * 3044 3143: gap of 100 bp
 * 3144 5889: contig of 2746 bp in length
 * 5890 5989: gap of 100 bp
 * 5990 8979: contig of 2990 bp in length
 * 8980 9079: gap of 100 bp
 * 9080 13674: contig of 4595 bp in length
 * 13675 13774: gap of 100 bp
 * 13775 18831: contig of 5057 bp in length
 * 18832 18931: gap of 100 bp
 * 18932 23526: contig of 4595 bp in length
 * 23527 23626: gap of 100 bp
 * 23627 27386: contig of 3760 bp in length
 * 27387 27486: gap of 100 bp
 * 27487 32572: contig of 5086 bp in length
 * 32573 32672: gap of 100 bp
 * 32673 38632: contig of 5960 bp in length
 * 38633 38732: gap of 100 bp
 * 38733 43735: contig of 5003 bp in length
 * 43736 43835: gap of 100 bp
 * 43836 49020: contig of 5185 bp in length
 * 49021 49120: gap of 100 bp
 * 49121 53660: contig of 4540 bp in length
 * 53661 53760: gap of 100 bp
 * 53761 59544: contig of 5784 bp in length
 * 59545 59644: gap of 100 bp


```

Rattus
1 (bases 1 to 1491)
Icard-Liepkalns, C., Mallet, J. and Corresponding, N.A.
AUTHORS
Patent: WO 9827406-A 1 25-JUN-1998;
JOURNAL
ICARD LIEPKALNS CHRISTINE (FR); MALLETT JACQUES (FR)
FEATURES
Location/Qualifiers
Source
1..1491
/organism="Rattus sp."
/db_xref="taxon:10118"
/feature="unnamed protein product"
/protein_id="CAB69414.1"
/db_xref="GI:6740203"
/translation="MAPPLDAPTIQVSOEQTPFGASDHEVLSNSTPPSPTLVPR
DCSAEAGDCGTGRKURARRPKSELALSKORRRKKANDRRNRHNLNSA
LDALRGVLPDPDAKLTETLFAHNYIWAQTOTLRADHSFYGPVPPVPCGELGS
PGGSSGDWGSYSPVQAGSLPTASLEEFPLQVPSPSCLLPGLTVFSDFL"
BASE COUNT 307 a 487 c 413 g 284 t
ORIGIN
1..1491
US-09-595-947C-10 (1-214) x A91167 (1-1491)
Alignment Scores:
Pred. No.: 5,67e-33 Length: 1491
Score: 849.00 Matches: 166
Percent Similarity: 83.72% Conservative: 14
Best Local Similarity: 77.21% Mismatches: 33
Query Match: 75.33% Indels: 2
DB: 6 Gaps: 2
US-09-595-947C-10 (1-214) x A91167 (1-1491)
QY 1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
DB 459 ATGGCGCCTCATCCCTTGGATGCGCCACCATCCCAAGTGTCCAGAGACCCAGCAACCC 518
QY 21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40
DB 519 TTTCGGGAGCCTCGGACCAAGTGTCTCAGTTCCTCAATCCACCCACTAGCCCACT 578
QY 41 ArgThrProGlyAsnCysAlaGluAlaGluGluGlyCysArgGlyAlaProArgLys 60
DB 579 CTCGTACCGAGGAGTGTCTCCGAAGCAGAGCAGTGTGCTGCGGAGGACATCGAGGAG 638
QY 61 LeuArgAlaArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80
DB 639 CTCGTGCGCGGCGGAGGCGCACAGCCCAAGAGGAGTGTGCTGAGCAAGCAG 698
QY 81 ArgArgSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMethHisAspLeuAsn 100
DB 699 CGACGAGCGCGGCGAAGGCCAACGACCGGAGCGCAACCGCATGCAACCTTAC 758
QY 101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAlaLysLeuThr 120
DB 759 TCCGCGCTGGATGCGCTGCGGTGTCTGCGCACCTTCCCGATGACGCCAACCTTACA 818
QY 121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleThrAlaLeuThrGlnThrLeuArg 140
DB 819 AGATCGAGACCTCGCTGCGCCCAACTACATTGGGCACCTGACTCAGACGCTGCGC 878
QY 141 IleAlaAspHisSerLeuThrAlaLeuGluProProAlaProHisCysGlyGluLeuGly 160
DB 879 ATAGCGGACCACAGCTTCTACGCGCCCGAGCCCTGTCGCC---TGTGGGAGCTGGA 935
QY 161 SerPro---GlyGlyProGlyAspThrPglySerLeuThrSerProValSerGlnAla 179
DB 936 AGCCGGGAGGGGCTCCAGCGGAGTGTGCTCTACTCTCTCTCTCTCTCTCTCTCTCTCT 995
QY 180 GlySerLeuSerProAlaAlaSerLeuGluArgProGlyLeuLeuGlyAlaThrSer 199
DB 996 GGTAGCTGAGGCCACAGCCTCATTTGGAGGATTCCTTGGCTGCGAGTGTGCGGCTGCTCC 1055
QY 200 SerAlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214
DB 1055

```

```

Db 1056 CCATCTGTCTCTCCCGGACCCCTGGTGTCTCAGACTTCTTG 1100
RESULT 6
RNRELAXT
LOCUS
DEFINITION
R.norvegicus mRNA for transcriptional regulator, Relax.
ACCESSION
Y10619
VERSION
Y10619.1 GI:2072737
KEYWORDS
Relax; transcriptional regulator.
SOURCE
Rattus norvegicus
ORGANISM
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 1491)
AUTHORS
Ravassard, P., Chatail, F., Mallet, J. and Icard-Liepkalns, C.
TITLE
Relax, a novel rat bHLH transcriptional regulator transiently
expressed in the ventricular proliferating zone of the developing
central nervous system
J. Neurosci. Res. 48 (2), 146-158 (1997)
JOURNAL
MEDLINE
97276390
PUBMED
9130143
REFERENCE
2 (bases 1 to 1491)
AUTHORS
Ravassard, P.
TITLE
Direct Submission
JOURNAL
Submitted (20-JAN-1997) P. Ravassard, CNRS UMR 9923, Bat. CERVI,
Hopital de la Pitie Salpetriere, 83 Bd. de l'Hopital, F-75013
Paris, FRANCE
FEATURES
Location/Qualifiers
Source
1..1491
/organism="Rattus norvegicus"
/strain="Wistar"
/db_xref="taxon:10116"
/clone_lib="RT-PCR"
459..1103
/function="transcriptional regulator"
/codon_start=1
/product="Relax"
/protein_id="CAA71630.1"
/db_xref="GI:2072738"
/db_xref="SPTREMBL:O08718"
/translation="MAPPLDAPTIQVSOEQTPFGASDHEVLSNSTPPSPTLVPR
DCSAEAGDCGTGRKURARRPKSELALSKORRRKKANDRRNRHNLNSA
LDALRGVLPDPDAKLTETLFAHNYIWAQTOTLRADHSFYGPVPPVPCGELGS
PGGSSGDWGSYSPVQAGSLPTASLEEFPLQVPSPSCLLPGLTVFSDFL"
BASE COUNT 307 a 487 c 413 g 284 t
ORIGIN
1..1491
US-09-595-947C-10 (1-214) x RNRELAXT (1-1491)
QY 1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
DB 459 ATGGCGCCTCATCCCTTGGATGCGCCACCATCCCAAGTGTCCAGAGACCCAGCAACCC 518
QY 21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40
DB 519 TTTCGGGAGCCTCGGACCAAGTGTCTCAGTTCCTCAATCCACCCACTAGCCCACT 578
QY 41 ArgThrProGlyAsnCysAlaGluAlaGluGluGlyCysArgGlyAlaProArgLys 60
DB 579 CTCGTACCGAGGAGTGTCTCCGAAGCAGAGCAGTGTGCTGCGGAGGACATCGAGGAG 638
QY 61 LeuArgAlaArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80
DB 639 CTCGTGCGCGGCGGAGGCGCACAGCCCAAGAGGAGTGTGCTGAGCAAGCAG 698

```


REFERENCE 2 (bases 1 to 1861)
 AUTHORS Schwitzgebel, V. and German, M.
 TITLE Direct Submission
 JOURNAL Submitted (26-MAR-2001) Hormone Research Institute, University of California San Francisco, 513 Parnassus Ave, HSW1090, San Francisco, CA 94145-0534, USA

FEATURES
 source 1. 1861
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 <1093. >1737
 /product="neurogenin 3"
 1093. .1737
 /codon_start=1
 /product="neurogenin 3"
 /protein_id="AAK50058.1"
 /db_xref="GI:13937129"
 /translation="MAPHLDALAIQVSPQPPFGASDEVLSNSTPSPPLIPR
 DCSAEVGDGCRGTRKLRARRGGRNPKSELALSKRRSRKRRKANDERNRMHLNSA
 LDALRGVLPFPDDAKLTETLFAHNYIWAQTQLRIADHSFYGPPEPPVPCGELGS
 PGGSGNDWGSIIYSPVQAGNLSSTASLEEFPGIQVPSSPSYLLPGALVFSDFL"

BASE COUNT 397 a 560 c 537 g 367 t

ORIGIN

Alignment Scores:
 Pred. No.: 1.91e-31 Length: 1861
 Score: 819.00 Matches: 163
 Percent Similarity: 82.33% Conservative: 14
 Best Local Similarity: 75.81% Mismatches: 36
 Query Match: 72.67% Indels: 2
 DB: 10 Gaps: 2

US-09-595-947C-10 (1-214) x AF364300 (1-1861)

Qy 1 MethThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
 Db 1093 ATGGCGCCTCATCCCTTGGATCGCTCACCATCCAAAGTTCCTCCAGAGACACAACT 1152
 Qy 21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProSerProThr 40
 Db 1153 TTTCGGGAGCCTCGGACGACGAAAGTGTCTCAATTCACCCCTAGCCCTAGCCCACT 1212
 Qy 41 ArgThrProGlyAsnCysAlaGluAlaGluGluGlyCysArgGlyAlaProArgLys 60
 Db 1213 CTCATCTAGGACTGTCTCGAAGCAGAAAGTGGTGTACTCGCGAGGACCTCGAGGAG 1272
 Qy 61 LeuArgAlaArgArgGlyCysArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80
 Db 1273 CTCGCGCCCGACGGCGGAGCGCAACAGCCCAAGAGCGAGTGGCCTCAGCAAAACAG 1332
 Qy 81 ArgArgSerArgArgLysAlaAsnAspArgGluArgAsnArgMethHisAspLeuAsn 100
 Db 1333 CGAAGACCGCGCGCAAGAAGCCCAATATCGGGAGCGCAATCGCATGCACAACTCAAC 1392
 Qy 101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThr 120
 Db 1393 TCGGCGCTGGATCGCTCGCGGTCTCTGCCACCTTCCCGATGAGCGCAAACTTACA 1452
 Qy 121 LysLeuGluThrLeuArgPheAlaHisAsnTyrlleThrAlaLeuThrGlnThrLeuArg 140
 Db 1453 AAGATCGAGACCTCGCTCGCCCAACATACATCTGGCCTGACTGACTGAGCGCTGCGC 1512
 Qy 141 IleAlaAspHisSerLeuTyrrAlaLeuGluProProAlaProHisCysGlyGluLeuGly 160
 Db 1513 ATAGCGGACACAGCTTCTATGGCCCGGAGCCCTGTGCCC---TGTGGAGAGCTGGGG 1569
 Qy 161 SerPro---GlyGlyProGlyAspTrpGlySerLeuTyrrSerProValSerGlnAla 179
 Db 1570 AGCCCGGAGGTGGCTCCACGGGAGCTGGGGCTCTATCTACTCCCGACTCTCCCAAGCG 1629
 Qy 180 GlySerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeuLeuGlyAlaThrSer 199
 Db 1630 GGTAACCTGACCCCGCCCTCATTTGGAGGAATTCCTCGCTGCGCTGCGCTGCGCCAGCTCC 1689

Qy 200 SerAlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214
 Db 1690 CCATCCTATCTGCTCCCGGAGCACTGGTGTTCAGACTTCTTG 1734

RESULT 9
 MMATH4B
 LOCUS MMATH4B 5567 bp DNA linear ROD 25-OCT-2000
 DEFINITION M.musculus MATH4B gene.
 ACCESSION Y09167
 VERSION Y09167.2 GI:11065737
 KEYWORDS bHLH; mammalian atonal homolog 4B; MATH4B; neurogenin 3.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 5567)
 AUTHORS Cau, E., Gradwohl, G., Fode, C. and Guillemot, F.
 TITLE Mash1 activates a cascade of bHLH regulators in olfactory neuron progenitors
 JOURNAL Development 124 (8), 1611-1621 (1997)
 MEDLINE 97261963
 PUBMED 9108377
 REFERENCE 2 (bases 1 to 5567)
 AUTHORS Jacquemin, P., Durvieux, S.M., Jensen, J., Godfraind, C., Gradwohl, G., Guillemot, F., Madsen, O.D., Carmeliet, P., Dewerchin, M., Collen, D., Rousseau, G.G. and Lemaigre, F.P.
 TITLE Transcription factor hepatocyte nuclear factor 6 regulates pancreatic endocrine cell differentiation and controls expression of the proendocrine gene ngn3
 JOURNAL Mol. Cell. Biol. 20 (12), 4445-4454 (2000)
 MEDLINE 20285449
 PUBMED 10825208
 REFERENCE 3 (bases 1 to 5567)
 AUTHORS Gradwohl, G.J.
 TITLE Direct Submission
 JOURNAL Submitted (04-NOV-1996) G.J. Gradwohl, IGBMC, CNRS-INSERM-Universite Louis Pasteur, BP163, C.U. de Strasbourg, F-67404 ILLKIRCH cedex, FRANCE
 REMARK Revised by [4]
 REFERENCE 4 (bases 1 to 5567)
 AUTHORS Lemaigre, F.P.
 TITLE Direct Submission
 JOURNAL Submitted (01-AUG-2000) Lemaigre F.P., Hormone and Metabolic Research Unit, Louvain University Medical School, Avenue Hippocrate 75, box 7529, Brussels 1200, BELGIUM
 COMMENT On Oct 31, 2000 this sequence version replaced gi:1666087.
 FEATURES
 source 1. 5567
 /organism="Mus musculus"
 /strain="SV129"
 /db_xref="taxon:10090"
 /clone="lambda GEM12 clone 19C"
 /cell_line="D3"
 /cell_type="embryonic stem cells"
 /clone_lib="Genomic library No.62, MboI partial digests cloned in BamHI"
 /dev_stage="embryos"
 958
 misc_binding /bound_moiety="HNF6"
 3692
 misc_binding /bound_moiety="HNF6"
 4145. .5567
 gene /gene="MATH4B"
 4145. .4151
 TATA_signal /gene="MATH4B"
 /note="putative"
 4923. 5567
 CDS /gene="MATH4B"
 /codon_start=1
 /product="neurogenin 3"
 /protein_id="CAA70366.1"
 /db_xref="GI:1666088"

/db_xref="SWISS-PROT:P70661"

/translation="NAPPLDALTIOVSPETQPPPGASDHVLSNTPPTLIPR
DCSEAEVGCCTCKLRRGCRNRKPKSELASLORSRRKKANDRENRMHNLNSA
LDALRGVTPDDAKLKIETLRFAHNYIWAQTQIRIADHSFDPPEVPVCGEIGS
PGGSGNGWSYISVPSQAGNISPLASUEEFGVQVPSYLLPGLVALISDFL"

BASE COUNT 1271 a 1549 c 1564 g 1183 t

ORIGIN

Alignment Scores:
Pred. No.: 5,09e-31 Length: 5567
Score: 819.00 Matches: 163
Percent Similarity: 82.33% Conservative: 14
Best Local Similarity: 75.81% Mismatches: 36
Query Match: 72.67% Indels: 2
DB: 10 Gaps: 2

US-09-595-947c-10 (1-214) x MMATH4B (1-5567)

QY 1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
DB 4923 ATGGCGCCTCATCCTTGGATGGCTCACCATCAAGTGTCGCCAGACACACAACCT 4982
QY 21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40
DB 4983 TTTCGCGGAGCCPCGACACGACGAGTGCTCAGTTCCAATTCACCCACCTAGCCCACT 5042
QY 41 ArgThrProGlyAsnGlyAlaGluGluGluGlyCysArgGlyAlaProArgLys 60
DB 5043 CTCATACCTAGGACGTCTCCGAGGAGAGAGTGGTGACTGCGAGGGACCTCGAGGAAG 5102
QY 61 LeuArgAlaArgArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80
DB 5103 CTCGCGCGCGACGCGAGGCGCAACAGCCAGAGCGAGTGTCGACTCAGCAACACAG 5162
QY 81 ArgArgSerArgArgLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn 100
DB 5163 CGAAGAAGCGCGGCAAGAGGCAATGATCGGAGGCGCAATCGCATGCACACCTCAAC 5222
QY 101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAlaLysLeuThr 120
DB 5223 TCGCGCTGGATCGCTGCGCGTGCTCCGCCACCTTCCGATGACGCGCAACATACA 5282
QY 121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArg 140
DB 5283 AAGATCGAGACCTGCGCTCGCCCACTACATCTGGCACTGAGCTGAGCGCTGCGC 5342
QY 141 IleAlaAspHisSerLeuTyrAlaLeuGluProAlaProHisCysGlyGluLeuGly 160
DB 5343 ATAGCGGACACAGCTTCTATGCGCGGAGCCCTGTGCCC---TGTGGAGAGCTGGGG 5399
QY 161 SerPro---GlyGlyProGlyAspTrpGlySerLeuTyrSerProValSerGlnAla 179
DB 5400 AGCCCGGAGGTGGCTCCACGCGGAGCTGGGGCTCTATCTATCCCGAGTCCCAAGCG 5459
QY 180 GlySerLeuSerProAlaAlaSerLeuGluArgProGlyLeuLeuGlyAlaThrSer 199
DB 5460 GGTAACTGAGCCCAACGCGCTCATTTGGAGGAATTCCTGCGCTGCGAGGTCGCGCTCC 5519
QY 200 SerAlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214
DB 5520 CCATCCTATCTGTCGCGGAGCACTGGTGTCTCTGACACTCTTG 5564

RESULT 10
AC109783/c
LOCUS AC109783 138070 bp DNA linear HTG 07-FEB-2002
DEFINITION Mus musculus clone RP23-121F10, WORKING DRAFT SEQUENCE, 17
unordered pieces.
ACCESSION AC109783
VERSION AC109783.1 GI:18581594
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Mus musculus.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 138070)

McCombie,W.R., Baker,J.P., Ballia,V., Dedhia,N.N., de la
Bastide,M., Katzenberger,F., Kuit,K., King,L., Kirchoff,K.A.,
Miller,B., Muller,S., Nascimento,L.U., O'Shaughnessy,A.L.,
Preston,R.R., Santos,L., Spiegel,L.A., Palmer,L., Yang,C. and
Zutavern,T.

Mouse Genomic Sequence

Unpublished

2 (bases 1 to 138070)

McCombie,W.R.

Direct Submission

Submitted (07-FEB-2002) Lita Annenberg Hazen Genome Sequencing
Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring
Harbor, NY 11724, USA

----- Genome Center -----

Center: Lita Annenberg Hazen Genome Center, Cold Spring Harbor

Laboratory

Center code: CSHL

Web site: <http://www.cshl.org/genseq>

Contact: mcombie@cshl.org

----- Project Information

Project name: RP23-121F10

Clone name: RP23-121F10

Insert size: 173000; agarose-fp

Quality coverage: 4.00 in Q20 bases; agarose-fp

Quality coverage: 3.70 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 17 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 17785: contig of 17785 bp in length
* 17786 17874: gap of unknown length
* 17875 30356: contig of 12482 bp in length
* 30357 30444: gap of unknown length
* 30445 42306: contig of 11862 bp in length
* 42307 42394: gap of unknown length
* 42395 53598: contig of 11204 bp in length
* 53599 64355: gap of unknown length
* 64356 64443: gap of unknown length
* 64444 74016: contig of 9573 bp in length
* 74017 83366: contig of 9262 bp in length
* 83367 83455: gap of unknown length
* 83456 92356: contig of 8901 bp in length
* 92357 92444: gap of unknown length
* 92445 100821: contig of 8378 bp in length
* 100822 100909: gap of unknown length
* 100910 107529: contig of 6620 bp in length
* 107530 107617: gap of unknown length
* 107618 114066: contig of 6449 bp in length
* 114067 114154: gap of unknown length
* 114155 118873: contig of 4719 bp in length
* 118874 118961: gap of unknown length
* 118962 123619: contig of 4658 bp in length
* 123620 123707: gap of unknown length
* 123708 128240: contig of 4533 bp in length
* 128241 128328: gap of unknown length
* 128329 132682: contig of 4354 bp in length
* 132683 132770: gap of unknown length
* 132771 136341: contig of 3571 bp in length
* 136342 136429: gap of unknown length
* 136430 138070: contig of 1641 bp in length.

Location/Qualifiers

1. 138070

/organism="Mus musculus"

FEATURES

source

/db_xref="taxon:10090"
/clone="RP23-121F10"

BASE COUNT 35731 a 33657 c 33303 g 33954 t 1425 others
ORIGIN

Alignment Scores: 8.97e-30 Length: 138070
Pred. No.: 819.00 Matches: 163
Score: 82.33% Conservative: 14
Best Local Similarity: 75.81% Mismatches: 36
Query Match: 72.67% Indels: 2
DB: 2 Gaps: 2

US-09-595-947c-10 (1-214) x AC109783 (1-138070)

QY 1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
Db 111220 ATGGCGCCTCATCCCTTGGATGGCTCACCATCCCAAGTGTCCCGAGACACAAACCT 111161
QY 21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40
Db 111160 TTTCGGGAGCCTCGGACGACCAAGTGTCTCAGTTCCAAATCCACCCACCTAGCCCACT 111101
QY 41 ArgThrProGlyAsnCysAlaGluAlaGluGluGlyGlyCysArgGlyAlaProArgLys 60
Db 111100 CTCATACCTAGGAGCTGCTCCGAAGCAGAGTGGGTGACTGCGGAGGACCTCGAGGAAG 111041
QY 61 LeuArgAlaArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80
Db 111040 CTCGCGCGCCGCGCGGAGGCGCACAGCCCAAGAGCGGTGGCACTCAGCAACACAG 110981
QY 81 ArgArgSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn 100
Db 110980 CGAAGAAGCGCGCGCAAGAAGCGCAATGATCGGGAGCGCAATGCGATGCAACACCTCAAC 110921
QY 101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAlaLysLeuThr 120
Db 110920 TCGGCGCTGGATGCGTGGCGGTGTCTGCCACCTTCCGGATGACGCCCAACTACA 110861
QY 121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArg 140
Db 110860 AAGATCGAGACCTGCGCTTCCGCCACACTACATCTGGCACTGACTGAGACGCTGCGC 110801
QY 141 IleAlaAspHisSerLeuThrAlaLeuGluProProAlaProHisCysGlyGluLeuGly 160
Db 110800 ATAGCGGACACAGCTTCTATGCGCGGAGCCCTGTGCCCCTTCCGGATGACGCCCAACTACA 110744
QY 161 SerPro---GlyGlyProGlyAspTrpGlySerLeuTyrSerProValSerGlnAla 179
Db 110743 AGCCCGGAGGTGGCTCCACGGGAGTGGGGCTCTACTACTCCCACTCTCCCAAGCG 110684
QY 180 GlySerLeuSerProAlaLeuSerLeuGluGluArgProGlyLeuLeuGlyAlaThrSer 199
Db 110683 GGTAACTGAGCCCGGCGCTCATTTGGAGGAATTCCTGGCTGAGGTGCCAGCTCC 110624
QY 200 SerAlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214
Db 110623 CCATCCTATCTGCTCCGGGAGCAGTGGTGTCTCAGACTTCTG 110579

RESULT 11

AC127417 185806 bp DNA linear HTG 15-JUL-2002
LOCUS
DEFINITION Mus musculus chromosome UNK clone RP23-459M2, WORKING DRAFT
SEQUENCE, 36 unordered pieces.

AC127417

AC127417.1 GI:21759524

VERSION HTG: HTGS_PHASE1; HTGS_DRAFT.

KEYWORDS house mouse.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 185806)

AUTHORS McPherson, J.D. and Waterston, R.H.

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

The sequence of Mus musculus clone
Unpublished
2 (bases 1 to 185806)
McPherson, J.D. and Waterston, R.H.
Direct Submission
Submitted (15-JUL-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc/index.shtml>
Contact: submissions@wustl.edu
----- Project Information -----
Center project name: M_BA0459M02

----- Summary Statistics -----
Sequencing vector: M13; 0%

Chemistry: Dye-terminator Big Dye; 100%

Assembly: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990319

Consensus quality: 173518 bases at least Q40

Consensus quality: 177276 bases at least Q30

Consensus quality: 179611 bases at least Q20

Insert size: 192000; agarose-fp

Insert size: 184535; sum-of-contrigs

Quality coverage: 3.66 in Q20 bases; agarose-fp

Quality coverage: 3.85 in Q20 bases; sum-of-contrigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 36 contrigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contrigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1043: contig of 1042 bp in length
* 1042: contig of 1042 bp in length
* 1143: gap of unknown length
* 1142: gap of unknown length
* 2216: contig of 1074 bp in length
* 2316: gap of unknown length
* 2317: contig of 1522 bp in length
* 3838: contig of 1522 bp in length
* 3938: gap of unknown length
* 3939: contig of 1458 bp in length
* 5397: gap of unknown length
* 5497: contig of 1378 bp in length
* 6874: gap of unknown length
* 6974: gap of unknown length
* 6975: contig of 1507 bp in length
* 8481: gap of unknown length
* 8482: contig of 1293 bp in length
* 8582: gap of unknown length
* 8583: contig of 1509 bp in length
* 9975: gap of unknown length
* 11483: contig of 1509 bp in length
* 11484: gap of unknown length
* 11583: gap of unknown length
* 13834: contig of 2251 bp in length
* 13835: gap of unknown length
* 13935: contig of 2101 bp in length
* 16035: gap of unknown length
* 16135: contig of 2965 bp in length
* 19100: gap of unknown length
* 19101: contig of 2984 bp in length
* 22184: gap of unknown length
* 22185: contig of 3003 bp in length
* 25287: gap of unknown length
* 25288: contig of 3044 bp in length
* 25388: gap of unknown length
* 28432: contig of 2878 bp in length
* 28433: gap of unknown length
* 31409: contig of 2878 bp in length
* 31410: gap of unknown length
* 35320: contig of 3811 bp in length
* 35321: gap of unknown length
* 38808: contig of 3388 bp in length
* 38809: gap of unknown length
* 43846: contig of 4938 bp in length

```

* 43847 43946: gap of unknown length
* 43947 49171: contig of 5225 bp in length
* 49172 49971: gap of unknown length
* 49272 53157: contig of 3886 bp in length
* 53158 53257: gap of unknown length
* 53258 57731: contig of 4474 bp in length
* 57732 57831: gap of unknown length
* 57832 62660: contig of 4829 bp in length
* 62661 68243: gap of unknown length
* 68244 68343: gap of unknown length
* 68344 73047: contig of 4704 bp in length
* 73048 73148: gap of unknown length
* 73149 78566: contig of 5418 bp in length
* 78567 85727: gap of unknown length
* 85728 91985: gap of unknown length
* 91986 92086: gap of unknown length
* 92087 97985: contig of 6159 bp in length
* 97986 105338: contig of 3899 bp in length
* 105339 105439: gap of unknown length
* 105440 113368: contig of 7254 bp in length
* 113369 113468: contig of 7930 bp in length
* 113469 120753: gap of unknown length
* 120754 120853: contig of 7285 bp in length
* 120854 129342: gap of unknown length
* 129343 129443: gap of unknown length
* 129444 138998: contig of 9556 bp in length
* 138999 139099: gap of unknown length
* 139100 149580: contig of 10481 bp in length
* 149581 149680: gap of unknown length
* 149681 162343: contig of 12664 bp in length
* 162344 162444: gap of unknown length
* 162445 185806: contig of 23363 bp in length.

```

FEATURES

source

```

1. 185806
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="UNK"
/clone="RP23-459M2"
1. 1042
/misc_feature 1. 1042
/misc_feature 1143. 2216
/misc_feature 2317. 3838
/misc_feature 3939. 5396
/misc_feature 5497. 6874
/misc_feature 6975. 8481
/misc_feature 8582. 9874
/misc_feature 9975. 11483
/misc_feature 11584. 13834
/misc_feature 13935. 16035
/misc_feature 16136. 19100
/misc_feature 19201. 22184
/misc_feature 22285. 25287
/misc_feature 25388. 28431
/misc_feature 28532. 31409
/misc_feature 31510. 35320

```

```

/misc_feature /note="assembly_name:Contig39"
35421. 38808
/misc_feature /note="assembly_name:Contig40"
38909. 43846
/misc_feature /note="assembly_name:Contig41"
43947. 49171
/misc_feature /note="assembly_name:Contig42"
49272. 53157
/misc_feature /note="assembly_name:Contig43"
53258. 57731
/misc_feature /note="assembly_name:Contig44"
57832. 62660
/misc_feature /note="assembly_name:Contig45"
62761. 68243
/misc_feature /note="assembly_name:Contig46"
68344. 73047
/misc_feature /note="assembly_name:Contig47"
73148. 78566
/misc_feature /note="assembly_name:Contig48"
78666. 85726
/misc_feature /note="assembly_name:Contig49"
85827. 91985
/misc_feature /note="assembly_name:Contig50"
92086. 97984
/misc_feature /note="assembly_name:Contig51"
vector_side:left
98085. 105338
/misc_feature /note="assembly_name:Contig52"
105439. 113368
/misc_feature /note="assembly_name:Contig53"
113469. 120753
/misc_feature /note="assembly_name:Contig54"
120854. 129342
/misc_feature /note="assembly_name:Contig55"
129443. 138998
/misc_feature /note="assembly_name:Contig56"
139099. 149579
/misc_feature /note="assembly_name:Contig57"
149680. 162343
/misc_feature /note="assembly_name:Contig58"
162444. 185806

```

Alignment Scores:

```

Pred. No.: 1.17e-29 Length: 185806
Score: 819.00 Matches: 163
Percent Similarity: 82.33% Conservative: 14
Best Local Similarity: 75.81% Mismatches: 36
Query Match: 72.67% Indels: 2
DB: 2 Gaps: 2

```

US-09-595-947c-10 (1-214) x AC127417 (1-185806)

```

QY 1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
Db 158824 ATGGCGCTCATCTGGATGCTCCATCAAGTGTCCCGACAGACACAAACCT 158883
QY 21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40
Db 158884 TTTCGGGAGCCCTCGGACACAGAGTGTCTCAGTTCCAAATCCACCCACCTAGCCCACT 158943
QY 41 ArgThrProGlyAsnCysAlaGluAlaGluGluGlyCysArgGlyAlaProArgLys 60
Db 158944 CTCATACCTAGGAGCTGCTCCGAAGCAGAAAGTGGGTGACTGCCGAGGAGCTCGAGGAAG 159003
QY 61 LeuArgAlaArgArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80
Db 159004 CTCGGCGCGGAGCGGAGGCGCAACAGGCCCAAGAGGAGTGGCAGCTCAGCAACAG 159063
QY 81 ArgArgSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn 100
Db 159064 CGAAGAGCGCGGCGCAAGAGGCCAATGATCGGAGCGCAATGCGATGCAACACCTCAAC 159123

```

```

QY 101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAlaLysLeuThr 120
Db 159124 TGGGCGCTGGATGGCTGCGGCTGTCCTCCACCTTCCCGATGACGCCAACTTACA 159183
QY 121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArg 140
Db 159184 AAGATCGAGACCGCTGGCTGCCACCACTACATCTGGCACTGACTCAGACGCTGCC 159243
QY 141 IleAlaAspHisSerLeuTyrAlaLeuGluProProAlaProHisCysGlyGluLeuGly 160
Db 159244 ATAGGGGACACAGCTTCATGGCCGGAGCCCTGTGCC---TGTGAGAGCTGGGG 159300
QY 161 SerPro---GlyGlyProGlyAspTrpGlySerLeuTyrSerProValSerGlnAla 179
Db 159301 ACCCCGGAGTGGCTCCACGGGAGCTGGGCTCTATCTCTCCCACTCTCCCAAGCG 159360
QY 180 GlySerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeuLeuGlyAlaThrSer 199
Db 159361 GGTAACCTGAGCCACCGGCTCATGGAGGAATCCCTGGCTGCGAGTGCCAGCTCC 159420
QY 200 SerAlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214
Db 159421 CCATCCTATCTGCTCCGGGAGCACTGGTGTCTTCAGACTTCTTG 159465

RESULT 12
AC011010/c
LOCUS AC011010 170896 bp DNA linear HTG 16-MAR-2000
DEFINITION Homo sapiens clone RP11-6P16, WORKING DRAFT SEQUENCE, 21 unordered
pieces.
ACCESSION AC011010
VERSION AC011010.4 GI:7107881
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.

REFERENCE
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., DeArelano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrum,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
JOURNAL Direct Submission
COMMENT Submitted (29-SEP-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Feb 28, 2000 this sequence version replaced gi:6479051.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L2916
Center clone name: 6_P_16
----- Summary Statistics
Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads

```

```

Assembly program: Phrap; version 0.960731
Consensus quality: 114103 bases at least Q40
Consensus quality: 141555 bases at least Q30
Consensus quality: 158230 bases at least Q20
Insert size: 154000; agarose-fp
Quality coverage: 3.6 in Q20 bases; agarose-fp
Quality coverage: 3.3 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 21 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1744: contig of 1744 bp in length
* 1745 1844: gap of 100 bp
* 1845 2954: contig of 1110 bp in length
* 2955 3054: gap of 100 bp
* 3055 4839: contig of 1785 bp in length
* 4840 4939: gap of 100 bp
* 4940 7461: contig of 2522 bp in length
* 7462 7561: gap of 100 bp
* 7562 11912: contig of 4351 bp in length
* 11913 12012: gap of 100 bp
* 12013 15127: contig of 3115 bp in length
* 15128 15227: gap of 100 bp
* 15228 19176: contig of 3949 bp in length
* 19177 19276: gap of 100 bp
* 19277 24838: contig of 5562 bp in length
* 24839 24938: gap of 100 bp
* 24939 30632: contig of 5694 bp in length
* 30633 30732: gap of 100 bp
* 30733 37821: contig of 7089 bp in length
* 37822 37921: gap of 100 bp
* 37922 44638: contig of 6717 bp in length
* 44639 44738: gap of 100 bp
* 44739 49424: contig of 4686 bp in length
* 49425 49524: gap of 100 bp
* 49525 56517: contig of 6993 bp in length
* 56518 56617: gap of 100 bp
* 56618 65413: contig of 8796 bp in length
* 65414 65513: gap of 100 bp
* 65514 75659: contig of 10146 bp in length
* 75660 75759: gap of 100 bp
* 75760 86433: contig of 10674 bp in length
* 86434 86533: gap of 100 bp
* 86534 98763: contig of 12230 bp in length
* 98764 98863: gap of 100 bp
* 98864 109919: contig of 11056 bp in length
* 109920 110019: gap of 100 bp
* 110020 125264: contig of 15245 bp in length
* 125265 125364: gap of 100 bp
* 125365 141043: contig of 15679 bp in length
* 141044 141143: gap of 100 bp
* 141144 170896: contig of 29753 bp in length.
FEATURES
source
1..170896
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="RP11-6P16"
/clone_lib="RP11-6P16"
1..1744
/note="assembly_fragment"
1845..2954
/note="assembly_fragment"
3055..4839
/note="assembly_fragment"
4940..7461
/note="assembly_fragment"
7562..11912
misc_feature
misc_feature
misc_feature
misc_feature
misc_feature

```

misc_feature /note="assembly_fragment"
 12013..15127
 /note="assembly_fragment"
 15228..19176
 /note="assembly_fragment"
 19277..24838
 /note="assembly_fragment"
 24939..30632
 /note="assembly_fragment"
 30733..37821
 /note="assembly_fragment"
 37922..44638
 /note="assembly_fragment"
 44739..49424
 /note="assembly_fragment"
 49525..58517
 /note="assembly_fragment"
 56618..65413
 /note="assembly_fragment"
 65514..75659
 /note="assembly_fragment"
 75760..88433
 /note="assembly_fragment"
 clone_end:SP6
 vector_side:left
 86534..98763
 /note="assembly_fragment"
 98864..109919
 /note="assembly_fragment"
 110020..125264
 /note="assembly_fragment"
 125365..141043
 /note="assembly_fragment"
 141144..170896
 /note="assembly_fragment"
 41547 a 41714 c 41241 g 44377 t 2017 others

Alignment Scores:
 Pred. No.: 1.8e-15 Length: 170896
 Score: 523.00 Matches: 99
 Percent Similarity: 98.02% Conservative: 0
 Best Local Similarity: 98.02% Mismatches: 2
 Query Match: 46.41% Indels: 0
 DB: 2 Gaps: 0

US-09-595-947C-10 (1-214) x AC011010 (1-170896)

QY 114 ProAspAlaLysLeuThrLysIleGluThrLeuArgPheAlaHisAsnTyrIleTrp 133
 DB 30631 CCACGACGCGGAAGCTCACCAGATCGAGACGCTGCGCTTCGCCCACTACATCTGG 30572
 QY 134 AlaLeuThrGlnThrLeuArgIleAlaAspHisSerLeuTyrAlaLeuGluProProAla 153
 DB 30571 GCGTGACTCAACGCTGCGCATAGCGACACACAGCTTGACGCGCTGGAGCGCGCGG 30512
 QY 154 ProHisCysGlyGluLeuGlySerProGlyGlyProProGlyAspTrpGlySerLeuTyr 173
 DB 30511 CCGCAGCTGCGGGAGCTGGGCGAGCCCGGCGTCCCGGGGACTGGGGTCCCTCTAC 30452
 QY 174 SerProValSerGlnAlaGlySerLeuSerProAlaAlaSerLeuGluArgProGly 193
 DB 30451 TCCCCAGTCTCCAGGCTGGCAGCTGAGTCCCGCGCTGCTGGAGGACCGCGG 30392
 QY 194 LeuLeuGlyAlaThrSerSerAlaCysLeuSerProGlySerLeuAlaPheSerAspPhe 213
 DB 30391 CTGCTGGGGGCGACCTTTTCGCTGCTTGAGCCCGAGCTGCTGCTTCTCAGATTTT 30332
 QY 214 Leu 214
 DB 30331 CTG 30329
 RESULT 13

GGA012659 790 bp DNA linear VRT 03-JAN-2001
 LOCUS Gallus gallus ngn2/ath4a gene.
 DEFINITION AJ012659
 VERSION AJ012659.1 GI:3892740
 KEYWORDS atonal protein; bHLH transcription factor; neurogenin.
 SOURCE chicken.
 ORGANISM Gallus gallus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
 Phasianinae; Gallus.
 REFERENCE 1 (bases 1 to 790)
 AUTHORS Matter-Sadzinski, L., Matter, J. M., Ong, M. T., Hernandez, J. and Ballivet, M.
 TITLE Specification of neurotransmitter receptor identity in developing retina: the chick ARH5 promoter integrates the positive and negative effects of several bHLH proteins
 JOURNAL Development 128 (2), 217-231 (2001)
 MEDLINE 21064448
 PUBMED 11124117
 REFERENCE 2 (bases 1 to 790)
 AUTHORS Ballivet, M.
 TITLE Direct Submission
 JOURNAL Submitted (16-NOV-1998) Ballivet M., Biochemistry, University of Geneva, 30, quai Ernest Ansermet, 1211 Geneva 4, SWITZERLAND
 FEATURES
 source
 1..790
 /organism="Gallus gallus"
 /db_xref="taxon:9031"
 /clone="p56.1"
 /clone.lib="lambda L47, homemade"
 116..757
 /gene="ngn2"
 /note="ath4a"
 116..757
 /gene="ngn2"
 /function="bHLH transcription factor"
 /note="neurogenin 2"
 /codon_start=1
 /evidence=experimental
 /product="atonal-like protein 4a"
 /protein_id="CAA10105.1"
 /db_xref="GI:3892741"
 /db_xref="SPTREMBL:O9YH82"
 /translation="MPVKASPPAAEDLELLRLASPPASASLPSSAGBDEDEEDG
 RPRRLQEGARRRQRPRAARTAAETAKRRRLKANNRRNRNHNNAALDLR
 DVLPTFPEDAKLTKIETLRFHNYIWALETLLRAGARLGGADAAAPGAAAGSPSP
 ASSWSGASPPASPPVACTLSPGSPAGSADAEHWPGRFAPPPPHRL"
 BASE COUNT 91 a 351 c 283 g 65 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 4.28e-11 Length: 790
 Score: 388.50 Matches: 104
 Percent Similarity: 54.31% Conservative: 22
 Best Local Similarity: 44.83% Mismatches: 63
 Query Match: 34.47% Indels: 43
 DB: 5 Gaps: 10
 US-09-595-947C-10 (1-214) x GGA012659 (1-790)
 QY 3 ProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSerPhePro 22
 DB 47 CCACGCGCCCGCAGCCGCCGACGCGCC---GTCACACGAGCGCGCGCCCTCTCTCCC 103
 QY 23 Arg-----AlaSerGluAspGluVal 29
 DB 104 CGCCCCCGCAGGATGCGGTGAAGCGGAGAGCCGCCGCCGCGGAGGAGCACTG 163
 QY 30 ThrCysProThrSerAlaProProSerProThrArgThr----ProGlyAsnCysAlaGlu 48
 DB 164 CTGCTGTGCGCTCGCGTCCGCCGCCCTCGCGCTCGCTGCGCTCGCGCGCGGAG 223

SOURCE
ORGANISM
Norway rat.
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE
AUTHORS
1 (bases 1 to 105572)
Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C.,
Alsbrooks, S.L., Amaratunga, H.C., Are, J.R., Avele, M., Banks, T.,
Barbaria, J., Benton, J., Bimage, K., Blankenburg, K., Bonnin, D.,
Bouck, J., Bowie, S., Brivea, M., Brown, E., Brown, M., Bryant, N.P.,
Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,
Carton, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,
Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,
Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,
Devila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,
Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,
Douthwater, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,
Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,
Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,
Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,
Herrandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B.,
Homs, J., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E.,
Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,
Karlssoon, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,
Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Louseged, H.,
Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,
Massey, E., Mawhinney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M.,
Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
Nguyen, N., Nickerson, E., Nwokukwu, S., Oguh, M., Okwuonu, G.,
Oragunye, N., Oriado, R., Pace, A., Payton, B., Peery, J., Perez, L.,
Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,
Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savary, G.,
Scherer, S., Scott, G., Shen, H., Shoohtari, N., Sisson, I.,
Sodergren, E., Sonaik, T., Sparks, A., Stanley, H., Stone, H.,
Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
Tansley, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q.,
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K.,
Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
Weinstock, G. and Gibbs, R.

TITLE
JOURNAL
Direct Submission
Unpublished
2 (bases 1 to 105572)
Worley, K.C.

REFERENCE
AUTHORS
TITLE
JOURNAL
Submitted (19-FEB-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 105572)
Worley, K.C.

REFERENCE
AUTHORS
TITLE
JOURNAL
Submitted (13-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 12, 2002 this sequence version replaced gi:18701957.

COMMENT
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GPAB
Center clone name: CH230-235K10
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329

Consensus quality: 42382 bases at least Q40
Consensus quality: 45708 bases at least Q30
Consensus quality: 49494 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 55 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1274: contig of 1274 bp in length
* 1375 1374: gap of unknown length
* 1375 2888: contig of 1514 bp in length
* 2889 2988: gap of unknown length
* 2989 4406: contig of 1418 bp in length
* 4407 4506: gap of unknown length
* 4507 6096: contig of 1590 bp in length
* 6097 6196: gap of unknown length
* 6197 7656: contig of 1460 bp in length
* 7657 7756: gap of unknown length
* 7757 9085: contig of 1329 bp in length
* 9086 9185: gap of unknown length
* 9186 10439: contig of 1244 bp in length
* 10439 10539: gap of unknown length
* 10539 11704: contig of 1175 bp in length
* 11705 11804: gap of unknown length
* 11805 13085: contig of 1281 bp in length
* 13086 13185: gap of unknown length
* 13186 14926: contig of 1741 bp in length
* 14927 15026: gap of unknown length
* 15027 16549: contig of 1523 bp in length
* 16550 16649: gap of unknown length
* 16650 18165: contig of 1516 bp in length
* 18166 18265: gap of unknown length
* 18266 19724: contig of 1459 bp in length
* 19725 19824: gap of unknown length
* 19825 20829: contig of 1005 bp in length
* 20830 20929: gap of unknown length
* 20930 22321: contig of 1392 bp in length
* 22322 22421: gap of unknown length
* 22422 23592: contig of 1171 bp in length
* 23593 23692: gap of unknown length
* 23693 25176: contig of 1484 bp in length
* 25177 25276: gap of unknown length
* 25277 27028: contig of 1752 bp in length
* 27029 27128: gap of unknown length
* 27129 28815: contig of 1687 bp in length
* 28816 28915: gap of unknown length
* 28916 30149: contig of 1234 bp in length
* 30150 30249: gap of unknown length
* 30250 31482: contig of 1233 bp in length
* 31483 31582: gap of unknown length
* 31583 33013: contig of 1431 bp in length
* 33014 33113: gap of unknown length
* 33114 34306: contig of 1193 bp in length
* 34307 34406: gap of unknown length
* 34407 35732: contig of 1326 bp in length
* 35733 35832: gap of unknown length
* 35833 36987: contig of 1155 bp in length
* 36988 37087: gap of unknown length
* 37088 38488: contig of 1401 bp in length
* 38489 38588: gap of unknown length
* 38589 40099: contig of 1511 bp in length
* 40100 40199: gap of unknown length
* 40200 41975: contig of 1776 bp in length
* 41976 42075: gap of unknown length
* 42076 43414: contig of 1339 bp in length
* 43415 43514: gap of unknown length
* 43515 45944: contig of 2430 bp in length

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 8, 2003, 21:14:33 ; Search time 252 Seconds
(without alignments)
1912.411 Million cell updates/sec

Title: US-09-595-947c-10

Perfect score: 1127

Sequence: 1 MTPQSGAPTQVTRERS.....LGATSSACLSPGSLAFSDFL 214

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2.1/uspto_spool/US09595947/runat_07042003_160404_20055/app_query.fasta_1.391
-DB=N_Geneseq_101002 -QFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptn -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09595947 -CGN_1_1_263 -runat_07042003_160404_20055 -NCPU=6 -ICPU=3
-NO_XLPHY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMECUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N_Geneseq_101002.*
1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1105	98.0	5340	21	AAC61089	Human neurogenin 3
2	849	75.3	1491	19	AAV42512	CDNA encoding a no
3	819	72.7	804	19	AAV27050	Mouse neurogenin 3
4	819	72.7	804	21	AAZ51981	Murine neurogenin-3
5	819	72.7	861	22	AAF27266	Murine neurogenin 3
6	819	72.7	1861	21	AAC61090	Murine neurogenin
7	819	72.7	5567	22	AAF27254	Mouse atonal homol
c 8	519	46.1	592	24	ABQ49522	Oligonucleotide fo
9	519	46.1	592	24	ABQ49523	Oligonucleotide fo
10	467	41.4	592	24	ABQ49524	Oligonucleotide fo
c 11	467	41.4	592	24	ABQ49525	Oligonucleotide fo
12	388.5	34.5	790	22	AAF27264	Chicken atonal hom
13	385.5	34.2	1074	22	AAF27263	Chicken atonal hom
14	378.5	33.6	1268	18	AAT74891	Human neurogenic d
15	378.5	33.6	1268	19	AAV42932	DNA encoding human
16	375.5	33.3	1385	19	AAV27049	Mouse neurogenin 2
17	375.5	33.3	1385	21	AAZ51980	Murine neurogenin
18	375.5	33.3	1385	22	AAF27269	Mouse neurogenin 2
19	370.5	32.9	1412	22	AAF27255	Mouse atonal homol
20	370.5	32.9	1412	22	AAF27273	Mouse atonal homol
21	370	32.8	1527	19	AAV27045	Rat neurogenin 1 9
22	370	32.8	1527	21	AAZ51976	Rat neurogenin-1 (
23	365	32.4	1332	19	AAV42938	DNA encoding murin
24	365	32.4	1333	18	AAT74894	Mouse neurogenic d
25	359.5	31.9	738	19	AAV27046	Murine neurogenin 1
26	359.5	31.9	738	21	AAZ51977	Murine neurogenin-
27	322.5	28.6	1312	19	AAV27047	Xenopus neurogenin
28	322.5	28.6	1312	21	AAZ51978	Xenopus X-ngnr-1a
29	304	27.0	1277	19	AAV27048	Xenopus neurogenin
30	304	27.0	1277	21	AAZ51979	Xenopus X-ngnr-1b
c 31	297.5	26.4	778	24	ABQ16590	Oligonucleotide fo
32	297.5	26.4	778	24	ABQ16591	Oligonucleotide fo
33	248	22.0	778	24	ABQ16592	Oligonucleotide fo
c 34	248	22.0	778	24	ABQ16593	Oligonucleotide fo
35	241	21.4	4161	23	ABL13239	Drosophila melanog
c 36	241	21.4	4161	23	ABL13238	Drosophila melanog
37	238.5	21.2	1550	22	AAF27276	Mouse atonal homol
38	238.5	21.2	1957	22	AAF27256	Mouse atonal homol
39	234.5	20.8	1344	21	AA60796	Human MATH-2 encod
40	234.5	20.8	1344	22	AAH76705	Human atonal prote
41	233.5	20.7	1588	22	AB82993	Human transcriptio
42	231.5	20.5	1550	21	AAA62681	Human NeuroD2 gene
43	230	20.4	1535	18	AAT74890	Human neurogenic d
44	230	20.4	1535	19	AAV42931	DNA encoding human
45	230	20.4	1951	18	AAT74893	Mouse neurogenic d

ALIGNMENTS

RESULT 1

AAC61089
ID AAC61089 standard; DNA; 5340 BP.
XX AAC61089;
AC AAC61089;
XX
DT 05-FEB-2001 (first entry)
XX Human neurogenin 3 (Ngn3) genomic DNA sequence.

Neurogenin 3; Ngn3; chromosome 10q22.1-22.2; cellular differentiation;
Islet cell precursor identification; diabetes mellitus; human; ds.

Homo sapiens.

Key Location/Qualifiers
CDS 3022..3666
FT /*tag= a

FT /product= "Ngn3"
 XX /note= "Neurogenin 3"
 PN W0200059936-A1.
 XX 12-OCT-2000.
 PD
 XX 28-MAR-2000; 2000WO-US08436.
 PF
 XX 06-APR-1999; 9905-0128180.
 PR (REGC) UNIV CALIFORNIA.
 XX German MS, Lin J;
 PA
 PI
 XX WPI; 2000-664989/64.
 DR P-PSDB; AAY85617.
 XX
 XX Novel human neurogenin 3 polypeptides and polynucleotides encoding
 PT them, useful for diagnosis, prevention and treatment of diabetes
 PT mellitus and to identify individuals at risk of diabetes -
 XX
 XX Claim 6; Page 46-48; 54pp; English.
 PS
 XX The human neurogenin 3 Ngn3 DNA sequence AAC61089 encodes the Ngn3
 CC protein AAY85617. The Ngn3 gene is located at chromosome position
 CC 10q22.1-22.2. The invention relates to the human Ngn3 nucleotide and
 CC protein sequences, and includes an antibody recognising the Ngn3 protein.
 CC Also included in the invention is a method for identifying an islet cell
 CC precursor, the method involves analysing a cell for the expression of the
 CC Ngn3 gene product, where detection of the product is indicative of an
 CC islet cell precursor. The Ngn3 DNA sequence is useful as a diagnostic
 CC reagent for detecting (in a subject) a predisposition to a defect in
 CC pancreatic islet cell function or formation associated with a defect in
 CC Ngn3 activity. The Ngn3 protein is useful for identifying beta-cell
 CC precursor cells expressing Ngn3, and to alter cellular differentiation in
 CC culture in vivo to produce new beta-cells to treat patients with diabetes
 CC mellitus.
 XX
 SQ Sequence 5340 BP; 1215 A; 1500 C; 1514 G; 1111 T; 0 other;

Alignment Scores:
 Pred. No.: 1.6e-59 Length: 5340
 Score: 1105.00 Matches: 211
 Percent Similarity: 99.07% Conservative: 1
 Best Local Similarity: 98.60% Mismatches: 2
 Query Match: 98.05% Indels: 0
 DB: 21 Gaps: 0

US-09-595-947c-10 (1-214) x AAC61089 (1-5340)

QY 1 MethThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
 |||
 DB 3022 ATACGCGCTCAACCTCGGTGGCCACCTGTCACAGTACCGCTGAGACGGAGCGGTCC 3081
 QY 21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40
 |||
 DB 3082 TTCCCCAGAGCCTCGGAAGACGAGTACCTGCCCGCCAGTCCCGCCCGCCAGCCCGCCACT 3141
 QY 41 ArgThrProGlyAsnCysAlaGluAlaGluGluGlyCysArgGlyAlaProArgLys 60
 |||
 DB 3142 CGCACAGGGGGAACCTCGCAGAGGCGGAAGAGGAGGCTGCCGAGGGGCCCGGAGGAG 3201
 QY 61 LeuArgAlaArgArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80
 |||
 DB 3202 CTCGGGCACGGCGGGGACGACGCGCGCTTAAGACGAGTGGCAGTGGCAGCAGCAG 3261
 QY 81 ArgArgSerArgArgLysLysAlaAspArgGluArgAsnArgMetHisAspLeuAsn 100
 |||
 DB 3262 CGAGGAGTGGCGAAAGAGGCCAACGCGGAGGCGCAATCGAATGACACACCTCAAC 3321
 QY 101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAlaLysLeuThr 120
 |||

DB 3322 TCGGCACTGACGCCCTGGCGGTGCTCTGCCACCTTCCACAGACGAGAGCTCACC 3381
 QY 121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArg 140
 |||
 DB 3382 AAGATCGAGACGCTGCGCTTCGCCACAACTAGATCTGGCGCTGACTCAACGCTGGC 3441
 QY 141 IleAlaAspHisSerLeuTyrAlaLeuGluProProAlaProHisCysGlyGluLeuGly 160
 |||
 DB 3442 ATAGCGACACACGCTTGTACGGCTGGAGCGCGCGCGCTCTACTCCAGCTCCAGGCTGGC 3501
 QY 161 SerProGlyGlyProProGlyAspTrpGlySerLeuTyrSerProValSerGlnAlaGly 180
 |||
 DB 3502 AGCCAGCGGTTCCCGCGGACTGGGGTCCCTCTACTCCAGCTCCAGGCTGGC 3561
 QY 181 SerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeuLeuGlyAlaThrSerSer 200
 |||
 DB 3562 AGCTGAGTCCCGCGCTGCTGGAGAGCGACCGCGGCTGTGGGGGCGCACCTCTTCC 3621
 QY 201 AlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214
 |||
 DB 3622 GCCTGCTTGGAGCCAGCAGTCTGGCTTCTCAGATTTCTG 3663

RESULT 2
 AAV42512
 ID AAV42512 standard; cDNA; 1491 BP.
 XX
 AC AAV42512;
 XX
 DT 05-OCT-1998 (first entry)
 XX
 DE cDNA encoding a novel BHLH protein designated RELAX.
 XX
 KW Basic helix-loop-helix; BHLH; RELAX; Rat Embryonic Longitudinal Axis;
 KW control; gene expression; transcriptional activator; targeting;
 KW protein expression; central nervous system; CNS; treatment;
 KW nervous system disorder; ss.
 XX
 OS Rattus sp.
 XX
 FH Key
 FT CDS 459..1103
 FT /*tag= a
 FT /product= RELAX
 XX
 PN W09827206-A2.
 XX
 PD 25-JUN-1998.
 XX
 PF 19-DEC-1997; 97WO-FR02368.
 XX
 PR 19-DEC-1996; 96FR-0015651.
 XX
 PA (RHON) RHONE-POULENC RORER SA.
 XX
 PI Mallet J, Ravassard P, Icard-Liepkalns C;
 XX
 DR WPI; 1998-362775/31.
 DR P-PSDB; AAW62991.
 XX
 PT Basic helix-loop-helix polypeptide and related nucleic acid - with
 PT transcriptional activity, for targeting expression of genes to
 PT central nervous system and treatment of nervous disease
 XX
 XX Claim 6; Page 20; 28pp; French.
 CC The present sequence encodes a basic helix-loop-helix (BHLH) type
 CC protein, designated RELAX (Rat Embryonic Longitudinal Axis) protein.
 CC The protein is used to control and participate in gene expression,
 CC by acting as a transcriptional activator, strictly dependent on the
 CC presence of an intact E box (CANNTG), particularly for targeting
 CC expression of proteins to the central nervous system (CNS). The
 CC nucleic acid sequence can be used to treat nervous system disorders,
 CC and antisense sequences can be used to control mRNA transcription.

```

XX SQ Sequence 1491 BP; 307 A; 487 C; 413 G; 284 T; 0 other;
Alignment Scores:
Pred. No.: 3,29e-44 Length: 1491
Score: 849.00 Matches: 166
Percent Similarity: 83.72% Conservative: 14
Best Local Similarity: 77.21% Mismatches: 33
Query Match: 75.33% Indels: 2
DB: 19 Gaps: 2

US-09-595-947C-10 (1-214) x AAV42512 (1-1491)
Qy 1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
Db 459 ATGGCGCTCATCTTGGATCGCCCAACCATCCCAAGTGTCCCAAGAGCAGCCAGCAACC 518
Qy 21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40
Db 519 TTTCCCGGAGCTCGGACCAAGAGTGTCTAGTTCATTCACCCACCTAGCCCACT 578
Qy 41 ArgThrProGlyAsnCysAlaGluAlaGluGluGlyCysArgGlyAlaProArgLys 60
Db 579 CTCGTACCGAGGAGCTGTCCGACGACAGCAGTGTCTGCGAGGAGCATCGAGGAAG 638
Qy 61 LeuArgAlaArgArgGlyCysArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80
Db 639 CTCCTGTGCGCGCGGAGGCGGCACACAGGCCCAAGCGAGTGTGGCACTAGCAAGCAG 698
Qy 81 ArgArgSerArgArgLysAlaAsnAspArgGluArgGlnArgMetHisAspLeuAsn 100
Db 699 CGACCAAGCGCGGCAAGAGCCCAACGACCGGAGCGCAACCGATGCACACCTTAAC 758
Qy 101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAlaLysLeuThr 120
Db 759 TCCGCGCTGATGCGCTGCGGCTGTCTGCGCACCTTCCCGATGACGCCAACTTACA 818
Qy 121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArg 140
Db 819 AAGATCGAGACCTCGCTTCCGCCCACTACATTTGGGCACTACTCAGACGCTGGCC 878
Qy 141 IleAlaAspHisSerLeuTyrAlaLeuGluProProAlaProHisCysGlyGluLeuGly 160
Db 879 ATAGCGGACCAAGCTTCTACGCGCCCGAGCCCTCTGTGCC---TGTGGGAGCTGGGA 935
Qy 161 SerPro---GlyCysProGlnProGlyAspTrpGlySerLeuTyrSerProValSerGlnAla 179
Db 936 AGCCCGGAGGAGGCGCTCCAGCGCGGAGTGGGCTCTATCTACTCCCAAGTTTCCCAAGCT 995
Qy 180 GlySerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeuLeuGlyAlaThrSer 199
Db 996 GGTACCTTGAGCCCAAGACCTCATGTAGGAGATTCCTTGGCTGCGAGTGCACGCTCC 1055
Qy 200 SerAlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214
Db 1056 CCATCTGTCTGCTCCGCGGCACTGCTGTGTTCTCAGACTTCTTG 1100

RESULT 3
AAV27050
ID AAV27050 standard; cDNA; 804 BP.
XX AC AAV27050;
XX DT 17-SEP-1998 (first entry)
XX DE Mouse neurogenin 3 gene.
XX KW ds; Mouse; neurogenin; expression vector; recombinant protein;
XX KW antibody; neurogenesis.
XX OS Mus sp.
XX FH Key Location/Qualifiers

```

```

FT CDS 160..804
FT /*tag= a
FT /product= "Mouse neurogenin 3"
XX WO9813491-A2.
XX 02-APR-1998.
XX 24-SEP-1997; 97WO-US17048.
XX 17-SEP-1997; 97US-0932411.
XX 27-SEP-1996; 96US-0722570.
XX 12-NOV-1996; 96US-0030864.
XX 19-DEC-1996; 96US-0772009.
XX (CALY ) CALIFORNIA INST OF TECHNOLOGY.
XX Anderson DJ, Ma Q, Sommer L;
XX WPI: 1998-230702/20.
XX P-PSDB; AAW54947.
XX Mouse neurogenins, useful in neurogenesis - and recombinant nucleic
XX acids and proteins derived from rat and Xenopus
XX Disclosure; Fig 9; 106pp; English.
XX The Mouse neurogenin 3 is one of several neurogenin proteins discussed
XX in the present invention. The neurogenin nucleic acids can be expressed
XX in a host cell, transformed using an expression vector, to produce
XX recombinant proteins. The proteins and the antibodies raised against
XX the proteins are useful in the study of neurogenesis.
XX SQ Sequence 804 BP; 171 A; 263 C; 225 G; 145 T; 0 other;

Alignment Scores:
Pred. No.: 1.27e-42 Length: 804
Score: 819.00 Matches: 163
Percent Similarity: 82.31% Conservative: 14
Best Local Similarity: 75.81% Mismatches: 36
Query Match: 72.67% Indels: 2
DB: 19 Gaps: 2

US-09-595-947C-10 (1-214) x AAV27050 (1-804)
Qy 1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
Db 160 ATGGCGCTCATCTTGGATCGCCCAACCATCCCAAGTGTCCCGAGAGACACAAACCT 219
Qy 21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40
Db 220 TTTCCCGGAGCTCGGACCAAGAGTGTCTAGTTCATTCACCCCACTAGCCCACT 279
Qy 41 ArgThrProGlyAsnCysAlaGluAlaGluGluGlyCysArgGlyAlaProArgLys 60
Db 280 CTCATACCTAGGAGCTGTCTCCGAAGCAGAAGTGGGTGACTGCGAGGAGGAGCTCGAGGAAG 339
Qy 61 LeuArgAlaArgArgGlyCysArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80
Db 340 CTCGCGCGCGGAGGCGGCACAGGCCCAAGCGAGTGTGGCACTGACCAACAG 399
Qy 81 ArgArgSerArgArgLysAlaAsnAspArgGluArgGlnArgMetHisAspLeuAsn 100
Db 400 CGAAGAGCGCGGCAAGAGCCCAATGATCGGAGGAGCAATCGGATGCACACCTCAAC 459
Qy 101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAlaLysLeuThr 120
Db 460 TCGGCGCTGGATGCGCTGCGGCTGTCTGCGCACCTTCCCGATGACGCCAACTTACA 519
Qy 121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArg 140
Db 520 AAGATCGAGACCTGCGCTTCCGCCCACTACTCTGGGCACTGACTCAGACGCTGGCC 579

```

QY 141 IleAlaAspHisSerLeuTyrAlaLeuGluProAlaProHisCysGlyGluLeuGly 160
 DB 580 ATACGGACACACATTCATGGCCGGAGCCCTGTGCC---TGTGGAGAGCTGGG 636
 QY 161 SerPro---GlyGlyProGlyAspTrpGlySerLeuTyrSerProValSerGlnAla 179
 DB 637 AGCCCGGAGTGGCTCCACGGGAGCTGGGCTCTATCTACTCCCGAGTCTCCCAAGG 696
 QY 180 GlySerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeuLeuGlyAlaThrSer 199
 DB 697 GGTAACTGAGCCCGCCGCTCATTTGGAGAAATTCCTCGCTGCGAGTGCCAGCTCC 756
 QY 200 SerAlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214
 DB 757 CCATCTCTATCTGCTCCCGGAGCACTGGTGTCTCTCAGACTTCTTG 801

RESULT 4
 AAZ51981
 ID AAZ51981 standard; DNA; 804 BP.
 AC AAZ51981;
 XX
 DT 04-JUL-2000 (first entry)
 XX
 DE Murine neurogenin-3 (NGN3), nucleic acid sequence.
 XX
 KW Neurogenin-3; NGN-3; non-neuronal cell; NNC; neurogenesis;
 KW Phox2a protein; neuronal subtype-specific marker; growth factor;
 KW neural differentiation; transplantation; neuronal dysfunction;
 KW optical nerve damage; auditory nerve damage, neurodegenerative disorder;
 KW neuroprotective; nontropic; anticonvulsant; antiparkinsonian; vulnary;
 KW cerebroprotective; immunosuppressant; antiinfectious; ds.
 XX
 OS Mus sp.
 XX
 FH Key Location/Qualifiers
 FT CDS 160..804
 FT /*tag= a
 FT /product= "Murine neurogenin-3 protein"
 XX
 PN WO200009676-A2.
 XX
 PD 24-FEB-2000.
 XX
 PF 13-AUG-1999; 99WO-US18525.
 XX
 PR 14-AUG-1998; 98US-0096630.
 XX
 PA (CALY) CALIFORNIA INST OF TECHNOLOGY.
 XX
 PI Anderson DJ, Lo L;
 XX
 WPI: 2000-256250/22.
 DR P-PSDB; AA70570.
 XX
 PT Inducing non-neuronal cells to differentiate into neurons and for
 PT non-neuronal cells to express a neuronal subtype-specific marker,
 PT comprising contacting the non-neuronal cells with a vector containing
 PT neurogenin nucleic acid -
 XX
 PS Claim 1; Fig 1J; 76pp; English.
 XX
 CC The patent discloses a method for inducing non-neuronal cells (NNC) to
 CC differentiate into neurons and for NNCs to express a neuronal subtype
 CC -specific marker. Transformed host cells are used as sources of neuronal
 CC and other growth factors; in culture for screening compounds that
 CC modulate neural differentiation or as sources of recombinantly produced
 CC neurogenins and Phox2a proteins for use in transplantation. The cells
 CC also have a variety of in vivo uses, e.g. for transplantation at sites of
 CC neuronal dysfunction e.g. patients with hearing or vision loss due to
 CC optical or auditory nerve damage, brain or spinal cord injuries, and
 CC neurodegenerative disorders e.g. Alzheimer's disease. The present
 CC sequence encodes murine neurogenin-3 (NGN-3), a transcription factor.

CC NNCs differentiate into neurons through the recombinant expression of a
 CC transcription factor that induces a core program of neurogenesis. Forced
 CC expression of murine NGN3 can elicit expression of at least some neuronal
 CC phenotypic markers even in NNCs.

XX SQ Sequence 804 BP; 171 A; 263 C; 225 G; 145 T; 0 other;

Alignment Scores:

Pred. No.: 1,27e-42 Length: 804
 Score: 819.00 Matches: 163
 Percent Similarity: 82.33% Conservative: 14
 Best Local Similarity: 75.81% Mismatches: 36
 Query Match: 72.67% Indels: 2
 DB: 21 Gaps: 2

US-09-595-947C-10 (1-214) x AAZ51981 (1-804)

QY 1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
 DB 160 ATGGCGCTCATCCTTGGATGGCTCACCATCCAAAGTGTCCCGAGAGACACACACCT 219
 QY 21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40
 DB 220 TTTCCCGGAGCTCGGACACGAGTGTCTCAGTTCCAAATCCACCCACCTAGCCCACT 279
 QY 41 ArgThrProGlyAsnCysAlaGluAlaGluGluGlyCysArgGlyAlaProArgLys 60
 DB 280 CTATACCTAGGACTGCTCCGAGGAGAGAGTGGTGACTGCCGAGGAGCTCGAGGAG 339
 QY 61 LeuArgAlaArgArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80
 DB 340 CTCGCGCGGACGCGGAGGCGCAACAGGCCAGGAGCGAGTGGCACTCAGCAACAG 399
 QY 81 ArgArgSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn 100
 DB 400 CGAAGAAGCGCGGCAAGAAGGCAATGATCGGAGGCGCAATGCGATGCAACCTCAAC 459
 QY 101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAlaLysLeuThr 120
 DB 460 TCGGCGCTGGATGGCTGCGGCTGCTCCGCCACCTCCCGGATGACGCGCAACTTACA 519
 QY 121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArg 140
 DB 520 AAGATGAGACCTGGCTTCGCCCACTACATCTGGCACTGAGTCACTGAGAGCTGCGC 579
 QY 141 IleAlaAspHisSerLeuTyrAlaLeuGluProProAlaProHisCysGlyGluLeuGly 160
 DB 580 ATAGCGGACACAGCTTCTATGCGCGGAGCCCTGTGCC---TGTGGAGAGCTGGG 636
 QY 161 SerPro---GlyGlyProGlyAspTrpGlySerLeuTyrSerProValSerGlnAla 179
 DB 637 AGCCCGGAGTGGCTCCACGGGAGCTGGGCTCTATCTACTCCCGAGTCTCCCAAGG 696
 QY 180 GlySerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeuLeuGlyAlaThrSer 199
 DB 697 GGTAACTGAGCCCGCCGCTCATTTGGAGAAATTCCTCGCTGCGAGTGCCAGCTCC 756
 QY 200 SerAlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214
 DB 757 CCATCTCTATCTGCTCCCGGAGCACTGGTGTCTCTCAGACTTCTTG 801

RESULT 5

AAF27266
 ID AAF27266 standard; cDNA; 861 BP.

XX AAF27266;

AC AAF27266;

XX 24-APR-2001 (first entry)

XX Mouse neurogenin 3 (ngn3) cDNA, SEQ ID NO:24.

XX Atonal; homologue; orthologue; atonal-associated protein; deafness;
 KW hearing impairment; vestibular effect; balance disorder; osteoarthritis;

KW cellular proliferation; cerebellar granule neuron; gene therapy;
 KW mechanoreceptive cell growth; auditory; osteopathic; cytostatic;
 KW transgenic animal; ss.
 XX
 OS Mus musculus.
 XX WO200073764-A2.
 XX
 PD 07-DEC-2000.
 XX
 PF 01-JUN-2000; 2000WO-US15410.
 XX
 PR 01-JUN-1999; 99US-0137060.
 XX
 PR 19-JAN-2000; 2000US-0176993.
 XX
 PA (BAYU) BAYLOR COLLEGE MEDICINE.
 XX
 XX Zoghbi HY, Bellen H, Birmingham N, Hassan B, Ben-Arie N;
 PI WPI: 2001-032190/04.
 DR P-PSDB; AAB60359.
 XX
 PT Therapeutic use of atonal-associated nucleic acids or amino acids, or
 PT any of its homologs or orthologs, for the treatment of e.g. deafness,
 PT osteoarthritis and abnormal cell proliferation.
 XX
 PS Disclosure; Page -: 142pp; English.
 XX
 CC The invention relates to the use of atonal-associated nucleic acid or
 CC amino acid sequence, or any of its homologues or orthologues as
 CC therapeutic agents for the treatment of deafness, partial hearing loss,
 CC vestibular effects due to damage or loss of inner hair cells,
 CC osteoarthritis and abnormal cell proliferation. The invention also
 CC encompasses methods of screening for compounds which affect the
 CC expression of an atonal-associated nucleic acid sequence in an animal,
 CC and a transgenic animal in which an allele of a native atonal-associated
 CC gene is replaced by a heterologous nucleic acid sequence, thus
 CC inactivating the atonal-associated allele. The nucleic acids or proteins
 CC may be used in a method of treating an animal for hearing impairment,
 CC joint disease, balance disorders, abnormal cell proliferation, or other
 CC disease related to loss of a functional atonal-associated nucleic acid or
 CC protein. They may particularly be used to treat an animal with a
 CC deficiency in cerebellar granule neurons or their precursors, and may
 CC also be used in promoting mechanoreceptive cell growth and generating
 CC hair cells. The present sequence represents an atonal-associated nucleic
 CC acid sequence referred to in the invention.
 CC Note: The present sequence is not shown in the specification, but
 CC was obtained from GenBank.
 XX
 SQ Sequence 861 BP; 182 A; 274 C; 250 G; 155 T; 0 other;

Alignment Scores:
 Pred. No.: 1.36e-42 Length: 861
 Score: 819.00 Matches: 163
 Percent Similarity: 82.33% Conservative: 14
 Best Local Similarity: 75.81% Mismatches: 36
 Query Match: 72.67% Indels: 2
 DB: 22 Gaps: 2

US-09-595-947C-10 (1-214) x AAF27266 (1-861)

QY 1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrArgSer 20
 Db 160 ATGGCGCTCATCCCTGGATGGCTCACCACCAAGTGTCCCGAGACACAAACACCT 219
 QY 21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProSerProThr 40
 Db 220 TTTCGGGAGCTCGAGCCAGCCAGGCTCAGTTCACATTCACCCACCTAGCCCACT 279
 QY 41 ArgThrProGlyAsnGlyAlaGluGluGluGlyGlyCysArgGlyAlaProArgLys 60
 Db 280 CTCATACCTAGGACTGCTCCGAGGAGAGAGTGGGTGACTGCCGAGGACCTCGAGGAG 339

QY 61 LeuArgAlaArgArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80
 Db 340 CTCGCCGCCGACGCGGAGGCGCAACAGGCCCAAGAGGAGTTGGCACTCAGCAACAG 399
 QY 81 ArgArgSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMethHisAspLeuAsn 100
 Db 400 CGAAGAGCGCGCGCAAGAAGCCCAATGATCGGAGCGCAATCGCATGCACAACTCAAC 459
 QY 101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThr 120
 Db 460 TCGGGCGTGGATCGCTGCGCGTGTCTCCGCCACCTTCCCGGATGACGCCAACTTACA 519
 QY 121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArg 140
 Db 520 AAGATCGAGACCCCTGGCTTCGCCCACTACATCTGGGCACTGACTCAGACGCTGGCG 579
 QY 141 IleAlaAspHisSerLeuTyrAlaLeuGluProProAlaPheHisCysGlyGluLeuGly 160
 Db 580 ATAGCGGACCAACAGCTTCTATGCGCGGAGCGCCCTGTGCC---TGTGGAGAGCTGGG 636
 QY 161 SerPro---GlyGlyProProGlyAspTrpGlySerLeuTyrSerProValSerGlnAla 179
 Db 637 AGCCCGGAGGTGGCTCCAAACGGGAGCTGGGCTCTATCTACTCCCACTCTCCCAAGCG 696
 QY 180 GlySerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeuLeuGlyAlaThrSer 199
 Db 697 GGTAACCTGAGCCCGACGGCTCATTTGGAGGAATTCCTCCCTGGCTGCGAGTGGCCACTCC 756
 QY 200 SerAlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214
 Db 757 CCATCTCTCTCTGCTCCCGGAGCAGCTGGTGTCTCAGACTTCTTG 801
 RESULT 6
 AAC61090
 ID AAC61090 standard; DNA; 1861 BP.
 XX
 AC AAC61090;
 XX
 DT 05-FEB-2001 (first entry)
 XX
 DE Murine neurogenin 3 (Ngn3) genomic DNA sequence.
 XX
 KW Neurogenin 3; Ngn3; cellular differentiation; diabetes mellitus;
 KW islet cell precursor identification; mouse; ds.
 XX
 OS Mus musculus.
 XX
 FH Key Location/Qualifiers
 FT CDS 1093..1737
 FT /tag= a
 FT /product= "Ngn3"
 FT /note= "Neurogenin 3"
 XX
 XX WO200059936-A1.
 PN 12-OCT-2000.
 XX
 PD 28-MAR-2000; 2000WO-US08436.
 XX
 PF 06-APR-1999; 99US-0128180.
 XX
 PR (REGC) UNIV CALIFORNIA.
 PA German MS, Lin J;
 XX
 PI WPI: 2000-664989/64.
 XX
 DR P-PSDB; AAY85618.
 XX
 PT Novel human neurogenin 3 polypeptides and polynucleotides encoding
 PT them, useful for diagnosis, prevention and treatment of diabetes
 PT mellitus and to identify individuals at risk of diabetes.
 XX
 PS Claim 18; Page 49-50; 54pp; English.

XX The human neurogenin 3 Ngn3 DNA sequence AAC61089 encodes the Ngn3
 CC protein AAY85617. The Ngn3 gene is located at chromosome position
 CC 10q22.1-22.2. The invention relates to the human Ngn3 nucleotide and
 CC protein sequences, and includes an antibody recognising the Ngn3 protein.
 CC Also included in the invention is a method for identifying an islet cell
 CC precursor, the method involves analysing a cell for the expression of the
 CC Ngn3 gene product, where detection of the product is indicative of an
 CC islet cell precursor. The Ngn3 DNA sequence is useful as a diagnostic
 CC reagent for detecting (in a subject) a predisposition to a defect in
 CC pancreatic islet cell function or formation associated with a defect in
 CC Ngn3 activity. The Ngn3 protein is useful for identifying beta-cell
 CC precursor cells expressing Ngn3, and to alter cellular differentiation in
 CC culture in vivo to produce new beta-cells to treat patients with diabetes
 CC mellitus. The present sequence represents the murine Ngn3 genomic DNA
 CC sequence.

XX SQ Sequence 1861 BP; 397 A; 560 C; 537 G; 367 T; 0 other;

Alignment Scores:

Pred. No.: 3e-42 Length: 1861
 Score: 819.00 Matches: 163
 Percent Similarity: 82.33% Conservative: 14
 Best Local Similarity: 75.81% Mismatches: 36
 Query Match: 72.67% Indels: 2
 DB: 21 Gaps: 2

US-09-595-947C-10 (1-214) x AAC61090 (1-1861)

QY 1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
 Db 1093 ATGGCGCTCATCCCTGGATGCTGCTCACCATCCAGTGTCCCCAGACACAAACCT 1152
 QY 21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40
 Db 1153 TTTCGGCGCTCGGACCGACAGTGTCTAGTTCCAAATCCACCCACCTAGCCCACT 1212
 QY 41 ArgThrProGlyAsnCysAlaGluAlaGluGluGlyGlyCysArgGlyAlaProArgLys 60
 Db 1213 CTCATACCTAGGAGCTCTCCGAAGCAGAAAGTGGTGACTCCGAGGACCTCGAGGAG 1272
 QY 61 LeuArgAlaArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80
 Db 1273 CTCGCGCCCGACGCGAGGCGCGAACAGGCCCAAGAGCGAGTTGGCACTCAGCAACAG 1332
 QY 81 ArgArgSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn 100
 Db 1333 CGAAGAACCGCGCCAGAGAGCCCAATGATCGGAGCGCAATCCATCGACCACTCAAC 1392
 QY 101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAlaLysLeuThr 120
 Db 1393 TCGCGCTGTGATGCTGCGGTGTCTCCGCCACCTTCCGGATGACGCGCAAACTTACA 1452
 QY 121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArg 140
 Db 1453 AAGATCAGACCTGCTGCGCCCACTACATCTGGGCACTGATCAGACCTCGCG 1512
 QY 141 IleAlaAspHisSerLeuTyrAlaLeuGluProAlaProHisCysGlyGluLeuGly 160
 Db 1513 ATAGCGGACCACTTCTATGCGCGAGCCCTCTGCGCC---TCTGGAGAGCTGGG 1569
 QY 161 SerPro---GlyGlyProGlyAspTyrGlySerLeuTyrSerProValSerGlnAla 179
 Db 1570 AGCCCGGAGTGGCTCCACGCGGAGTGGGGCTCTATCTACTCCCGAGTCTCCCAAGCG 1629
 QY 180 GlySerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeuLeuGlyAlaThrSer 199
 Db 1630 GGTAACCTGAGCCCGCCGCTCATTTGGAGAAATTCCTTGGCTGAGGTTGCCAGCTCC 1689
 QY 200 SerAlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214
 Db 1690 CCATCTATCTGCTCCCGGAGGAGCTGGTGTCTTCAGACTTCTTG 1734

RESULT 7

AAF27254
 ID AAF27254 standard; CDNA; 5567 BP.

XX AAF27254;

XX 24-APR-2001 (first entry)

XX Mouse atonal homologue 5 (ATO5, Math4B) CDNA, SEQ ID NO:4.

XX Atonal; homologue; orthologue; atonal-associated protein; deafness;
 KW hearing impairment; vestibular effect; balance disorder; osteoarthritis;
 KW cellular proliferation; cerebellar granule neuron; gene therapy;
 KW mechanoreceptive cell growth; auditory; osteopathic; cytoskeletal;
 KW transgenic animal; ss.

XX Mus musculus.

XX WO200073764-A2.

XX 07-DEC-2000.

XX 01-JUN-2000; 2000WO-US15410.

XX 01-JUN-1999; 99US-0137060.

XX 19-JAN-2000; 2000US-0176993.

XX (BAYU) BAYLOR COLLEGE MEDICINE.

XX Zoghbi HY, Bellen H, Birmingham N, Hassan B, Ben-Arie N;

XX WPI; 2001-032190/04.

XX P-PSDB; AAB60350.

XX Therapeutic use of atonal-associated nucleic acids or amino acids, or

XX any of its homologs or orthologs, for the treatment of e.g. deafness,

XX osteoarthritis and abnormal cell proliferation -

XX Disclosure; Page -; 142pp; English.

XX The invention relates to the use of atonal-associated nucleic acid or
 CC amino acid sequence, or any of its homologs or orthologues as
 CC therapeutic agents for the treatment of deafness, partial hearing loss,
 CC vestibular effects due to damage or loss of inner hair cells,
 CC osteoarthritis and abnormal cell proliferation. The invention also
 CC encompasses methods of screening for compounds which affect the
 CC expression of an atonal-associated nucleic acid sequence in an animal,
 CC and a transgenic animal in which an allele of a native atonal-associated
 CC gene is replaced by a heterologous nucleic acid sequence, thus
 CC inactivating the atonal-associated allele. The nucleic acids or proteins
 CC may be used in a method of treating an animal for hearing impairment,
 CC joint disease, balance disorders, abnormal cell proliferation, or other
 CC disease related to loss of a functional atonal-associated nucleic acid or
 CC protein. They may particularly be used to treat an animal with a
 CC deficiency in cerebellar granule neurons or their precursors, and may
 CC also be used in promoting mechanoreceptive cell growth and generating
 CC hair cells. The present sequence represents an atonal-associated nucleic
 CC acid sequence referred to in the invention.

XX Note: The present sequence is not shown in the specification, but

XX was obtained from GenBank.

XX SQ Sequence 5567 BP; 1271 A; 1549 C; 1564 G; 1183 T; 0 other;

Alignment Scores:

Pred. No.: 9.25e-42 Length: 5567
 Score: 819.00 Matches: 163
 Percent Similarity: 82.33% Conservative: 14
 Best Local Similarity: 75.81% Mismatches: 36
 Query Match: 72.67% Indels: 2
 DB: 22 Gaps: 2

US-09-595-947C-10 (1-214) x AAF27254 (1-5567)

QY 1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
 DB 4923 ATGGCGCCTCATCCTTGATCGCTCACCATCCAGTGTCCCGAGACACAAACCT 4982
 QY 21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40
 DB 4983 TTTCGGGAGCCTCGGACCAAGGCTCAGTTCATTCACATTCACCCCACTAGCCCACT 5042
 QY 41 ArgThrProGlyAsnCysAlaGluAlaGluGlyGlyCysArgGlyAlaProArgLys 60
 DB 5043 CTCATACCTAGGAGTCTCCGAAGCAGAAAGTGGGTGACTCGGAGGACCTCGAGGAAG 5102
 QY 61 LeuArgAlaArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80
 DB 5103 CTCGGCGCCGACGGAGGGCGCAACAGCCCAAGAGGAGTGGCACTCAGCAACAG 5162
 QY 81 ArgArgSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMethHisAspLeuAsn 100
 DB 5163 CGAAGAAGCGCGGCAAGAAGCCATGATCGGGAGCGCAATCGCATGCACCACTCAAC 5222
 QY 101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAlaLysLeuThr 120
 DB 5223 TCGGGCTGGATCGCTGCGCGGTGCTTCCGCCACCTCCCGATGACGCCAACCTACA 5282
 QY 121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArg 140
 DB 5283 AGATCGAGACCTCGCTTCGCCACACACTACATCTGGCACTGACTCAGACGCTGCGC 5342
 QY 141 IleAlaAspHisSerLeuThrAlaLeuGluProProAlaProHisCysGlyGluLeuGly 160
 DB 5343 ATAGCGGACACAGCTTCTATGGCCCGGAGCCCTGTGCCC---TGTGGAGAGCTGGG 5399
 QY 161 SerPro---GlyGlyProGlyAspTrpGlySerLeuThrSerProValSerGlnAla 179
 DB 5400 AGCCCGGAGGTGGCTCCCAACGGGAGCTGGGCTCTATCTACTCCCACTCTCCCAAGCG 5459
 QY 180 GlySerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeuLeuGlyAlaThrSer 199
 DB 5460 GGTAACTAGCCCCACGGCTCATTTGGAGGAATTCCTGGCTGCGAGTGGCCCACTCC 5519
 QY 200 SerAlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214
 DB 5520 CCATCTATCTGCTCCGGGAGCACTGGTGTCTCTCAGACTCTTG 5564
 RESULT 8
 ID ABQ49522/c
 XX
 AC ABQ49522; standard; DNA; 592 BP.
 AC ABQ49522;
 DT 12-JUL-2002 (first entry)
 XX
 DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 36113.
 XX
 KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
 KW drug; side effect; cancer; central nervous system; cardiovascular;
 KW gastrointestinal; respiratory system; single nucleotide polymorphism;
 KW SNP; cell differentiation; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200218632-A2.
 XX
 PD 07-MAR-2002.
 XX
 PF 01-SEP-2001; 2001WO-BP10074.
 XX
 PR 01-SEP-2000; 2000DE-1043826.
 PR 05-SEP-2000; 2000DE-1044543.
 XX
 PA (EPIC-) EPIDENOMICS AG.
 XX
 PI Olek A, Piepenbrock C, Berlin K, Guetig D;

XX WPI; 2002-371829/40.
 DR
 XX
 PT Determining the degree of cytosine methylation in genomic DNA, useful
 PT for diagnosis and prognosis, comprises selective hybridization of
 PT amplicons from chemically treated DNA
 XX
 PS Claim 12; 56pp + Sequence Listing; 56pp; German.
 XX
 CC This invention describes a novel method for determining the degree of
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
 CC genomic sample of DNA. The sample is treated chemically to convert
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic
 CC DNA that contains the target C is amplified to form a labeled amplicon.
 CC The amplicon is hybridised to two classes, each with at least one
 CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
 CC and the degree of hybridisation to both classes is determined from the
 CC label on the amplicon. From the ratio of labels hybridised to the two
 CC classes of oligomers, the degree of methylation is calculated. The method
 CC is used: (i) for diagnosis and/or prognosis of side effects of
 CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
 CC of the central nervous, cardiovascular, gastrointestinal and respiratory
 CC systems etc., particularly by detecting mutations or single nucleotide
 CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
 CC types and for investigating cell differentiation. The method allows the
 CC methylation status of many C residues to be determined simultaneously.
 CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
 CC method for determining the degree of cytosine methylation described in
 CC the disclosure of the invention.
 XX
 SQ Sequence 592 BP; 81 A; 59 C; 201 G; 251 T; 0 other;
 Alignment Scores:
 Pred. No.: 3.8e-24 Length: 592
 Score: 519.00 Matches: 98
 Percent Similarity: 83.67% Conservative: 25
 Best Local Similarity: 66.67% Mismatches: 24
 Query Match: 46.05% Indels: 0
 DB: 24 Gaps: 0
 US-09-595-947C-10 (1-214) x ABQ49522 (1-592)
 QY 1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
 DB 443 ATAAACGCTCAACCTCGAATACGCTATCCAAATAACCCGTAACAAACGAAACGATCC 384
 QY 21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40
 DB 383 TTCCCAAAACCTCGAAACGAAATAACCTACCCACGCTCCGCCGCCCAACCCCACT 324
 QY 41 ArgThrProGlyAsnCysAlaGluAlaGluGlyGlyCysArgGlyAlaProArgLys 60
 DB 323 CGCACACGAAACAACTACGCAAAACGAAACAAACAAACAAACAAACAAACAAACAA 264
 QY 61 LeuArgAlaArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80
 DB 263 CTCGGAACAGCGCGCAAAACCGCAACCGACTTAAACACGAATTAACACATAACAAACAA 204
 QY 81 ArgArgSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMethHisAspLeuAsn 100
 DB 203 CGACGAAATCGACGAAACAAACCAACGACCGGAGCAATCGAATACACAACTCAAC 144
 QY 101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAlaLysLeuThr 120
 DB 143 TCGACACTAAACCGCTCAGCGATATCTTACCACCTTCCCAACGACGCGAACTCACC 84
 QY 121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArg 140
 DB 83 AAATCGAAACGCTAGCTTCGCCCACTACACTATCTAAACGCTAACTCAACGCTACGC 24
 QY 141 IleAlaAspHisSerLeuThr 147
 DB 23 ATAACGAAACCACTTATAC 3

```

RESULT 9
ABQ49523
ID ABQ49523 standard; DNA; 592 BP.
XX
AC ABQ49523;
XX
DT 12-JUL-2002 (first entry)
XX
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 36114.
XX
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.
XX
OS Homo sapiens.
XX
PN WO200218632-A2.
XX
PD 07-MAR-2002.
XX
PF 01-SEP-2001; 2001WO-EP10074.
XX
PR 01-SEP-2000; 2000DE-1043826.
XX
PS 05-SEP-2000; 2000DE-1044543.
XX
PA (EPiG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
DR WPI; 2002-371829/40.
XX
PT Determining the degree of cytosine methylation in genomic DNA, useful
PT for diagnosis and prognosis, comprises selective hybridization of
PT amplicons from chemically treated DNA
XX
PS Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
This invention describes a novel method for determining the degree of
methylation of a particular cytosine in a motif 5'-CpG-3', present in a
genomic sample of DNA. The sample is treated chemically to convert
cytosine (C) but not methylated C, to uracil, then part of the genomic
DNA that contains the target C is amplified to form a labeled amplicon.
The amplicon is hybridised to two classes, each with at least one
member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
and the degree of hybridisation to both classes is determined from the
label on the amplicon. From the ratio of labels hybridised to the two
classes of oligomers, the degree of methylation is calculated. The method
is used: (i) for diagnosis and/or prognosis of side effects of
therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
of the central nervous, cardiovascular, gastrointestinal and respiratory
systems etc., particularly by detecting mutations or single nucleotide
polymorphisms (SNPs); and (ii) for differentiation of cell or tissue
types and for investigating cell differentiation. The method allows the
methylation status of many C residues to be determined simultaneously.
ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
method for determining the degree of cytosine methylation described in
the disclosure of the invention.
XX
SQ Sequence 592 BP; 251 A; 201 C; 59 G; 81 T; 0 other;
Alignment Scores:
Pred. No.: 3.8e-24 Length: 592
Score: 519.00 Matches: 98
Percent Similarity: 83.67% Conservative: 25
Best Local Similarity: 66.67% Mismatches: 24
Query Match: 46.05% Indels: 0
DB: 24 Gaps: 0
US-09-595-947c-10 (1-214) x ABQ49523 (1-592)
Oy 1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20

```

```

:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 150 ATAAGCGCTCAACCTCGAATACGCCCACTATCAAAATAACCGGTAAACGAACGATCC 209
Oy 21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 210 TTCCCAAAACCTCGAAAACGAATAATACCTACCCACGTCGCCGCCGCCCAACCCACT 269
Oy 41 ArgThrProGlyAsnCysAlaGluAlaGluGlyGlyCysArgGlyAlaProArgLys 60
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 270 CGCACACGAAAACTACGCCAAAAACGAAAAAAAATACTACCGAAAAACCCCGAAAAA 329
Oy 61 LeuArgAlaArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 330 CTCGCAACACGACGCGAAAAACGCAACCGACTTAAAAACGAATTAACACTAAACAACAA 389
Oy 81 ArgArgSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn 100
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 390 CGACGAAATCGACGAAAAACCAACCAACGCGCAACCAATCGAATACACAACCTCAAC 449
Oy 101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAlaLysLeuThr 120
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 450 TCGACACTAAACGCCCTACGGGATATCTACCCACCTTCCCAACGACGGGAACTCACC 509
Oy 121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArg 140
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 510 AAAATCGAAACGCTACGCTTCGCCCAACACTACATCTAAACGCTAACTCAACGCTAGCG 569
Oy 141 IleAlaAspHisSerLeuTyr 147
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 570 ATAACGAACCAACTTATAC 590
RESULT 10
ABQ49524
ID ABQ49524 standard; DNA; 592 BP.
XX
AC ABQ49524;
XX
DT 12-JUL-2002 (first entry)
XX
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 36115.
XX
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.
XX
OS Homo sapiens.
XX
PN WO200218632-A2.
XX
PD 07-MAR-2002.
XX
PF 01-SEP-2001; 2001WO-EP10074.
XX
PR 01-SEP-2000; 2000DE-1043826.
XX
PS 05-SEP-2000; 2000DE-1044543.
XX
PA (EPiG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
DR WPI; 2002-371829/40.
XX
PT Determining the degree of cytosine methylation in genomic DNA, useful
PT for diagnosis and prognosis, comprises selective hybridization of
PT amplicons from chemically treated DNA
XX
PS Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
This invention describes a novel method for determining the degree of
methylation of a particular cytosine in a motif 5'-CpG-3', present in a
genomic sample of DNA. The sample is treated chemically to convert
cytosine (C) but not methylated C, to uracil, then part of the genomic
DNA that contains the target C is amplified to form a labeled amplicon.
The amplicon is hybridised to two classes, each with at least one
member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
and the degree of hybridisation to both classes is determined from the
label on the amplicon. From the ratio of labels hybridised to the two
classes of oligomers, the degree of methylation is calculated. The method
is used: (i) for diagnosis and/or prognosis of side effects of
therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
of the central nervous, cardiovascular, gastrointestinal and respiratory
systems etc., particularly by detecting mutations or single nucleotide
polymorphisms (SNPs); and (ii) for differentiation of cell or tissue
types and for investigating cell differentiation. The method allows the
methylation status of many C residues to be determined simultaneously.
ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
method for determining the degree of cytosine methylation described in
the disclosure of the invention.
XX
SQ Sequence 592 BP; 251 A; 201 C; 59 G; 81 T; 0 other;
Alignment Scores:
Pred. No.: 3.8e-24 Length: 592
Score: 519.00 Matches: 98
Percent Similarity: 83.67% Conservative: 25
Best Local Similarity: 66.67% Mismatches: 24
Query Match: 46.05% Indels: 0
DB: 24 Gaps: 0
US-09-595-947c-10 (1-214) x ABQ49523 (1-592)
Oy 1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20

```


CC DNA that contains the target C is amplified to form a labeled amplicon.
 CC The amplicon is hybridised to two classes, each with at least one
 CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
 CC and the degree of hybridisation to both classes is determined from the
 CC label on the amplicon. From the ratio of labels hybridised to the two
 CC classes of oligomers, the degree of methylation is calculated. The method
 CC is used: (i) for diagnosis and/or prognosis of side effects of
 CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
 CC of the central nervous, cardiovascular, gastrointestinal and respiratory
 CC systems etc., particularly by detecting mutations or single nucleotide
 CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
 CC types and for investigating cell differentiation. The method allows the
 CC methylation status of many C residues to be determined simultaneously.
 CC AB013410-AB04121 represent genomic DNA sequences used to illustrate the
 CC method for determining the degree of cytosine methylation described in
 CC the disclosure of the invention.

XX SQ Sequence 592 BP; 123 A; 59 C; 187 G; 223 T; 0 other;

Alignment Scores:
 Pred. No.: 6,4e-21 Length: 592
 Score: 467.00 Matches: 101
 Percent Similarity: 73.94% Conservativity: 4
 Best Local Similarity: 71.13% Mismatches: 37
 Query Match: 41.44% Indels: 0
 DB: 24 Gaps: 0

US-09-595-947C-10 (1-214) x ABQ49524 (1-592)

Qy 6 SerGlyAlaProThrValGlnValThrArgGluThrGluArgSerPheProArgAlaSer 25
 Db 165 TCGGTGGTGGTTTATTTAGTGGTTCGTGAGAGCGGTTTATTTAGAGTTTCG 224
 Qy 26 GluAspGluValThrCysProThrSerAlaProProSerProThrArgThrProGlyAsn 45
 Db 225 GAAGACGAAGTATTTTACGTTTCGTTTCGTTTATTTATTCGTACGGGGGAAT 284
 Qy 46 CysAlaGluAlaGluGluGlyCysArgGlyAlaProArgLysLeuArgAlaArg 65
 Db 285 TCGGTAGAGCGGAGAGAGAGGAGTTGTCGAGGGGTTTCGAGGAGTTTCGGTACGGCC 344
 Qy 66 GlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGlnArgArgSerArg 85
 Db 345 GGGGACGTTAGTCGTTTAAAGACGAGTTGTTAGTAACTAGCAGGAGTTCGGCA 404
 Qy 86 LysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsnSerAlaLeuAspAla 105
 Db 405 AAGAAGGTTAAGCATCGGAGCGTAATCGAATGTATATTTTAAATTCGGTATTGACGTT 464
 Qy 106 LeuArgGlyValLeuProThrPheProAspAlaLysLeuThrLysIleGluThrLeu 125
 Db 465 TTGCGCGGTTTGTGTTTATTTTATTTAGACGCGGAGTTTATTAAGATCGAGAGTTG 524
 Qy 126 ArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArgIleAlaAspHisser 145
 Db 525 CGTTTCGTTTAAATATATTTTGGCGTTGATTAAACGTTTCGTTATACGCGATTATAGT 584
 Qy 146 LeuTyr 147
 Db 585 TTGTAC 590

RESULT 11

ABQ49525/C

ID ABQ49525 standard; DNA; 592 BP.

XX AC ABQ49525;

XX DT 12-JUL-2002 (first entry)

XX DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 36116.

XX KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;

XX KW drug; side effect; cancer; central nervous system; cardiovascular;

KW gastrointestinal; respiratory system; single nucleotide polymorphism;
 KW SNP; cell differentiation; ds.

XX OS Homo sapiens.

XX PN WO200218632-A2.

XX PD 07-MAR-2002.

XX PF 01-SEP-2001; 2001WO-EP10074.

XX PR 01-SEP-2000; 2000DE-1043826.

XX PR 05-SEP-2000; 2000DE-1044543.

XX XX (EPIG-) EPIGENOMICS AG.

PI Olek A, Piepenbrock C, Berlin K, Guetig D;

XX WPI; 2002-371829/40.

XX Determining the degree of cytosine methylation in genomic DNA, useful
 PT for diagnosis and prognosis, comprises selective hybridization of
 PT amplicons from chemically treated DNA -

PS Claim 12; 56pp + Sequence Listing; 56pp; German.

XX This invention describes a novel method for determining the degree of
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
 CC genomic sample of DNA. The sample is treated chemically to convert
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic
 CC DNA that contains the target C is amplified to form a labeled amplicon.
 CC The amplicon is hybridised to two classes, each with at least one
 CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
 CC and the degree of hybridisation to both classes is determined from the
 CC label on the amplicon. From the ratio of labels hybridised to the two
 CC classes of oligomers, the degree of methylation is calculated. The method
 CC is used: (i) for diagnosis and/or prognosis of side effects of
 CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
 CC of the central nervous, cardiovascular, gastrointestinal and respiratory
 CC systems etc., particularly by detecting mutations or single nucleotide
 CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
 CC types and for investigating cell differentiation. The method allows the
 CC methylation status of many C residues to be determined simultaneously.
 CC AB013410-AB04121 represent genomic DNA sequences used to illustrate the
 CC method for determining the degree of cytosine methylation described in
 CC the disclosure of the invention.

XX SQ Sequence 592 BP; 223 A; 187 C; 59 G; 123 T; 0 other;

Alignment Scores:

Pred. No.: 6,4e-21 Length: 592
 Score: 467.00 Matches: 101
 Percent Similarity: 73.94% Conservativity: 4
 Best Local Similarity: 71.13% Mismatches: 37
 Query Match: 41.44% Indels: 0
 DB: 24 Gaps: 0

US-09-595-947C-10 (1-214) x ABQ49525 (1-592)

Qy 6 SerGlyAlaProThrValGlnValThrArgGluThrGluArgSerPheProArgAlaSer 25
 Db 428 TCGGGTGGTGGTTTATTTAGTGGTTCGTGAGAGCGGTTTATTTATAGAGTTTCG 369
 Qy 26 GluAspGluValThrCysProThrSerAlaProProSerProThrArgThrProGlyAsn 45
 Db 368 GAAGACGAAGTATTTTACGTTTCGTTTCGTTTATTTATTCGTACGGGGGAAT 309
 Qy 46 CysAlaGluAlaGluGluGlyCysArgGlyAlaProArgLysLeuArgAlaArg 65
 Db 308 TCGGTAGAGCGGAGAGAGAGGAGTTGTCGAGGGGTTTCGAGGAGTTTCGGTACGGCC 249
 Qy 66 GlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGlnArgArgSerArg 85
 Db 405 AAGAAGGTTAAGCATCGGAGCGTAATCGAATGTATATTTTAAATTCGGTATTGACGTT 464

Db 248 GGGGACGTAGTCGGTTTAAAGCGAGTGGTGTAGTAAGTAGGACGAGCTCGCGGA 189
 QY 86 LysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsnSerAlaLeuAspAla 105
 Db 188 AAGAGGTTAAGCATCGGAGCGATCAATGATGATATATATTTTAAATTCGGTATGGACGTT 129
 QY 106 LeuArgGlyValLeuProThrPheProAspAlaLysLeuThrLysIleGluThrLeu 125
 Db 128 TTGCGGGTGTGTTTATTTTATAGACGCGAGCTTTTATTATGATCGAGACGTTG 69
 QY 126 ArgPheAlaHisAsnTyrIleTrpAlaLeuThrGluThrLeuArgIleAlaAspHisSer 145
 Db 68 CGTTCGTTTATAATATATATTTGGCGGTTGATTAAACGTTCGGTATAGCGGATTATAGT 9
 QY 146 LeuTyr 147
 Db 8 TTGTAC 3
 RESULT 12
 AAF27264
 ID AAF27264 standard; cDNA; 790 BP.
 XX
 AC AAF27264;
 XX
 DT 24-APR-2001 (first entry)
 XX
 DE Chicken atonal homologue ngn2/ath4a cDNA, SEQ ID NO:20.
 XX
 KW Atonal; homologue; orthologue; atonal-associated protein; deafness;
 KW hearing impairment; vestibular effect; balance disorder; osteoarthritis;
 KW cellular proliferation; cerebellar granule neuron; gene therapy;
 KW mechanoreceptive cell growth; auditory; osteopathic; cytostatic;
 KW transgenic animal; ss.
 XX
 OS Gallus gallus.
 XX
 PN WO200073764-A2.
 XX
 PD 07-DEC-2000.
 XX
 PF 01-JUN-2000; 2000WO-US15410.
 XX
 PR 01-JUN-1999; 990US-0137060.
 PR 19-JAN-2000; 2000US-0176993.
 XX
 PA (BAYU) BAYLOR COLLEGE MEDICINE.
 XX
 PI Zoghbi HY, Bellen H, Birmingham N, Hassan B, Ben-Arie N;
 XX
 DR WPI; 2001-032190/04.
 DR P-PSDB; AAB60357.
 XX
 PT Therapeutic use of atonal-associated nucleic acids or amino acids, or
 PT any of its homologs or orthologs, for the treatment of e.g. deafness,
 PT osteoarthritis and abnormal cell proliferation -
 XX
 PS Disclosure; Page -; 142pp; English.
 XX
 CC The invention relates to the use of atonal-associated nucleic acid or
 CC amino acid sequence, or any of its homologues or orthologues as
 CC therapeutic agents for the treatment of deafness, partial hearing loss,
 CC vestibular effects due to damage or loss of inner hair cells,
 CC osteoarthritis and abnormal cell proliferation. The invention also
 CC encompasses methods of screening for compounds which affect the
 CC expression of an atonal-associated nucleic acid sequence in an animal,
 CC and a transgenic animal in which an allele of a native atonal-associated
 CC gene is replaced by a heterologous nucleic acid sequence, thus
 CC inactivating the atonal-associated allele. The nucleic acids or proteins
 CC may be used in a method of treating an animal for hearing impairment,
 CC joint disease, balance disorders, abnormal cell proliferation, or other
 CC disease related to loss of a functional atonal-associated nucleic acid or
 CC protein. They may particularly be used to treat an animal with a
 CC deficiency in cerebellar granule neurons or their precursors, and may

CC also be used in promoting mechanoreceptive cell growth and generating
 CC hair cells. The present sequence represents an atonal-associated nucleic
 CC acid sequence referred to in the invention.
 CC Note: The present sequence is not shown in the specification, but
 CC was obtained from GenBank.
 XX

SQ Sequence 790 BP; 91 A; 351 C; 283 G; 65 T; 0 other:

Alignment Scores:
 Pred. No.: 6,39e-16 Length: 790
 Score: 388.50 Matches: 104
 Percent Similarity: 54.31% Conservative: 22
 Best Local Similarity: 44.83% Mismatches: 63
 Query Match: 34.47% Indels: 43
 DB: 22 Gaps: 10

US-09-595-947C-10 (1-214) x AAF27264 (1-790)

QY 3 ProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSerPhePro 22
 Db 47 CCAGCGCCCCCGACCGCCGCGACGGCC---GTCGACCGCGCGCGCCCTCTCTCCC 103
 QY 23 Arg-----AlaSerGluAspGluVal 29
 Db 104 CGCCCCCGCAGGATGCGGTGAAGGGGAGAGCGCGCGCGCGCGGAGGAGCACTG 163
 QY 30 ThrCysProThrSerAlaProProSerProThrArgThr---ProGlyAsnCysAlaGlu 48
 Db 164 CTGCTGCTGCGCTCGCTCGCCCGCCCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCG 223
 QY 49 AlaGluGluGlyGlyCysArgGlyAlaProArgLysLeuArg-----AlaArgArgGly 66
 Db 224 GAGGACGAGGACGAGGAGGAGGCGCGCGCGCGCTGCGAGGCGCGCTCGCGGCG 283
 QY 67 GlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln-----Arg 81
 Db 284 GGGCGGACGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCATCAAG 343
 QY 82 ArgSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsnSer 101
 Db 344 CGGAGCGCGCGCTGAAGCCAAACCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 403
 QY 102 AlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAlaLysLeuThrLys 121
 Db 404 CGCTGGACGCGCTGCGCGACGTGCTGCCACCTTCCCGAGGAGCGCAAGCTCACCAAG 463
 QY 122 IleGluThrLeuArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArgIle 141
 Db 464 ATCGAGACGCTGCGCTTCGCCCAACTACATCTGGCGCTCACCGAGAGCTGCGCTG 523
 QY 142 AlaAspHisSer---LeuTyrAlaLeuGluProProAlaProHisCysGlyGluLeuGly 160
 Db 524 GCCGGGCGCGCGCTGGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 583
 QY 161 SerProGlyGlyProProGlyAspTrpGlySerLeuTyrSerProValSerGlnAlaGly 180
 Db 584 AGCCCC---TCGCGCGCGCTGCTGCTG-----AGCGGC 613
 QY 181 SerLeuSerProAlaAlaSerLeuGluArgProGlyLeuLeuGlyAlaThrSerSer 200
 Db 614 GCGCGCAGCGCGCGCGCGCTCC-----GCTCGCGCTAC 646
 QY 201 AlaCys---LeuSerProGlySerLeuAlaPheSer 211
 Db 647 GCTGCACCTTATCGCCCGCAGCGCGCGCGCGCTCC 682
 RESULT 13
 AAF27263
 ID AAF27263 standard; cDNA; 1074 BP.
 XX
 AC AAF27263;
 XX
 DT 24-APR-2001 (first entry)

XX Chicken atonal homologue ngn1/ath4c cDNA, SEQ ID NO:18.

DE Atonal; homologue; orthologue; atonal-associated protein; deafness;

XX hearing impairment; vestibular effect; balance disorder; osteoarthritis;

KW cellular proliferation; cerebellar granule neuron; gene therapy;

KW mechanoreceptive cell growth; auditory; osteopathic; cytostatic;

KW transgenic animal; ss.

XX Gallus gallus.

OS WO200073764-A2.

XX PN 07-DEC-2000.

XX PD 01-JUN-2000; 2000WO-US15410.

XX PF 01-JUN-1999; 99US-0137060.

XX PR 19-JAN-2000; 2000US-0176993.

XX (BAYU) BAYLOR COLLEGE MEDICINE.

XX Zoghbi HY, Bellen H, Birmingham N, Hassan B, Ben-Arie N;

PI WPI: 2001-032190/04.

XX DR P-PSDB; AAB60356.

XX Therapeutic use of atonal-associated nucleic acids or amino acids, or

PT any of its homologs or orthologs, for the treatment of e.g. deafness,

PT osteoarthritis and abnormal cell proliferation.

XX Disclosure; Page : 142pp; English.

XX The invention relates to the use of atonal-associated nucleic acid or

CC amino acid sequence, or any of its homologues or orthologues as

CC therapeutic agents for the treatment of deafness, partial hearing loss,

CC vestibular effects due to damage or loss of inner hair cells,

CC osteoarthritis and abnormal cell proliferation. The invention also

CC encompasses methods of screening for compounds which affect the

CC expression of an atonal-associated nucleic acid sequence in an animal,

CC and a transgenic animal in which an allele of a native atonal-associated

CC gene is replaced by a heterologous nucleic acid sequence, thus

CC inactivating the atonal-associated allele. The nucleic acids or proteins

CC may be used in a method of treating an animal for hearing impairment,

CC joint disease, balance disorders, abnormal cell proliferation, or other

CC disease related to loss of a functional atonal-associated nucleic acid or

CC protein. They may particularly be used to treat an animal with a

CC deficiency in cerebellar granule neurons or their precursors, and may

CC also be used in promoting mechanoreceptive cell growth and generating

CC hair cells. The present sequence represents an atonal-associated nucleic

CC acid sequence referred to in the invention.

CC Note: the present sequence is not shown in the specification, but

CC was obtained from GenBank.

XX Sequence 1074 Bp; 148 A; 449 C; 272 G; 205 T; 0 other;

SQ

Alignment Scores:

Pred. No.:	1,348-15	Length:	1074
Score:	385.50	Matches:	104
Percent Similarity:	51.29%	Conservative:	15
Best Local Similarity:	44.83%	Mismatches:	60
Query Match:	34.21%	Indels:	53
DB:	22	Gaps:	10

US-09-595-947c-10 (1-214) x AAF27263 (1-1074)

Qy 21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProSerPro--- 39

Db 228 TTTCCCTCCCTTTCCCTCCCTCC-----CCTCAGCACCTTCCCTCCCTCCCTCCCT 275

Qy 40 -----ThrArgThrProGlyAsnCysAlaGluAla 49

Db 276 GCGGCCCACTGAGCGGCTTCCTCCCGCAGGATGCTCCCGGAGCGGCGGAGCAGC 335

QY 50 GluGluGlyGlyCysArg-----GlyAlaProArgLysLeuArgAlaArgAGGly 66

Db 336 -----GGCGGGCTTTCGAGCGCGCGGAGCTCCCGGGAG-----CGGCGGAGGAGACGC 386

QY 67 GlyArgSerArgProLysSerGluLeuAlaLeuSerLysGlnArgArgSerArgArgLys 86

Db 387 GCGCGTGGCGGGCGGACCGAGGCTTTCCTGCACACCTCAACAGGAGCGCGCGGTG 446

QY 87 LysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsnSerLysAlaLeuAspAlaLeu 106

Db 447 AAAGCCAGCAGCGGAGCGGAACCGCATGCCACCTCAACCGCGCTGGATGAGCTC 506

QY 107 ArgGlyValLeuProThrPheProAspAlaLeuLysLeuThrLysLeuGluThrLeuArg 126

Db 507 CGCAGCGTCTCGCGACCTTCCCGGACGACACCAAACTCACCATAATCGAAACCTCGCGC 566

QY 127 PheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArgIleAlaAspHisSerLeu 146

Db 567 TTCGCTTACAACTACATCTGGGCCCTTCGAGAGACCTTCGTTTGGCGGAGCAGTGCCTC 626

QY 147 -----TyrAlaLeuGluProAlaProHisCysGlyGluLeuGlySerPro 162

Db 627 CCTCTCTCCCGCCCTTCGCGGGCGCGCGCGCC-----CCAGCGCC 671

QY 163 GlyGlyProGlyAspTrpGlySerLeuTyrSerPro-----Val 176

Db 672 GCGAGGAGCGCGGTTTCGTGGCTGCAGCGGTTCCCGCGCGCGCGCGCTCGCTCGCGCC 731

QY 177 SerGlnAlaGlySerLeuSerProAlaAlaSer----- 187

Db 732 TCGCGCTCGCGCGCGGACGAGCGCGCACCTCCGAGGACTCGGCTACGTCCCTCGGAC 791

QY 188 ---LeuGluGluArgProGlyLeuLeuGlyAlaThrSerSerAla----- 201

Db 792 GCGCTCGCGCGCTTCGCGGGCTGCGCGCGCGCGCGCGCGCGCGCTCCCTCGCGCTAGCCC 851

QY 202 -----CysLeuSerProGlySerLeuAlaPheSer 211

Db 852 TGCGCGTGGGTGTCTCGTCCCGCGCGCGCGCGCGCGCGCGCTCTCC 887

RESULT 14

AA74891

ID AA74891 standard; DNA; 1268 BP.

XX AA74891;

AC AA74891;

XX 02-OCT-1997 (first entry)

DT Human neurogenic differentiation protein (NeuroD3) DNA clone 20A1.

DE Neurogenic differentiation protein; NeuroD3 gene;

XX transcriptional activator; neuron; pancreas; gastrointestinal;

KW knock-out mouse; transgenic animal; cancer; diabetes; gene therapy;

XX ss.

XX Homo sapiens.

XX Key Location/Qualifiers

XX CDS 55..768

XX /*tag= a

XX WO9716548-A1.

XX 09-MAY-1997.

XX 30-OCT-1996; 96WO-US17532.

XX 02-NOV-1995; 95US-0552142.

XX (HUTC-) HUTCHINSON CANCER RES CENT FRED.

XX (WEIN/) WEINTRAUB N.

Score: 378.50 Matches: 103
Percent Similarity: 51.05% Conservative: 18
Best Local Similarity: 43.46% Mismatches: 63
Query Match: 33.58% Indels: 54
DB: 19 Gaps: 8

US-09-595-947C-10 (1-214) x AAV42932 (1-1268)

QY 4 GlnProSer-----GlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
DB 59 CAGCCCGCTTGAGACCTGCATCTCCGACCTGCGACTGCGCCAGCAGCAGCGCAGTGACC 118
QY 21 PheProArgAlaSerGluAspGluValThrCys-----ProThr 33
DB 119 TATCCGGCTTCTCACC--GACGAGGAGACTGTGCGACACTCCACAGCAGCCTCCGCT 177
QY 34 SerAlaProProSerProThrArgThrProGlyAsnCysAlaGluAlaGluGlyGly 53
DB 178 TCGGGCGCGCGCGCGCGCGCGC----- 201
QY 54 CysArgGlyAlaProArgLysLeuArgAla----- 63
DB 202 ---AGGGCGCGCCCAATATCTCCGGCGCTGTAGGTTCCAGGGGCACAGGACGACGAG 258
QY 64 -----ArgArgGlyArgSerArgProLysSerGluLeuAlaLeuSerLys 79
DB 259 CAGGAGAGCGCGCGCGCGCGCGCGCGCGCTCCGAGCGCTGCTGCACCTCG 318
QY 80 GlnArgArgSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMethHisAspLeu 99
DB 319 CTGCGCAGGAGCGCGCGCTCAAGCCACGATCGCGAGCGCAACCGCATGCACAACCTG 378
QY 100 AsnSerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAlaLysLeu 119
DB 379 AACGGCGCGCTGACGACACTGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 438
QY 120 ThrLysIleGluThrLeuArgPheAlaHisAsnTyrIleTyrAlaLeuThrGlnThrLeu 139
DB 439 ACCAAATCGACAGCGTGGCTTCGGCTTACAACTACATCTGGGCTCTGGCGGACACTG 498
QY 140 ArgIleAlaAspHisSerLeu-----TyrAlaLeuGluProPro 152
DB 499 CGCCTGGCGGATCAAGGGCTGCGCGGAGCGGTGCGCGGAGCGCTCTGCGCGCGCAG 558
QY 153 AlaProHisCysGlyGluLeuGlySerProGlyGlyProProGlyAsp-----Tyr 169
DB 559 TCGTCCCTTCG-----CTGCCCGGTCCCAAGCCCGCGGAGCGCGGAGTCTGG 612
QY 170 GlySer-----LeuTyrSerProValSerGlnAlaGlySerLeuSerProAlaAla 186
DB 613 GGCTCAGGTGCGCGCGCGCTCCCGCTCTGACCCAGTAGCCCGCGCTCCGAA 672
QY 187 SerLeuGluGluArgProGly-LeuLeuGlyAlaThrSerSerAlaCys 202
DB 673 GACTTCACCTACCGCGCGCGCGCGCTGTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 721

Search completed: April 8, 2003, 23:23:13
Job time : 260 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model
Run on: April 8, 2003, 23:15:09 ; Search time 564 Seconds
(without alignments)
332.826 Million cell updates/sec

Title: US-09-595-947C-10
Perfect score: 1127
Sequence: 1 MTPQSGATVQVTRERS.....LGATSSACLSPGSLAFSDFL 214

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 593429 seqs, 438593890 residues

Total number of hits satisfying chosen parameters: 1186858

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US09595947/runat_07042003_160406_20125/app_query.fasta_1.391
-DB=Published Applications_NA -OFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCI=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blom62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09595947 -ECGN_1_1_93=runat_07042003_160406_20125
-NCPU=6 -ICPU=3 -NO_XLPXY -NO_MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications_NA:
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1105	98.0	5340	10	US-09-817-360-1
2	819	72.7	645	9	US-10-004-717-4
3	819	72.7	861	9	US-10-004-717-24
4	819	72.7	1861	10	US-09-817-360-3

5	388.5	34.5	790	9	US-10-004-717-20
6	385.5	34.2	1074	9	US-10-004-717-18
7	375.5	33.3	1385	9	US-10-004-717-30
8	370.5	32.9	1412	9	US-10-004-717-6
9	370.5	32.9	1412	9	US-10-004-717-37
10	370	32.8	1527	7	US-08-722-570-12
11	359.5	31.9	738	7	US-08-722-570-13
12	322.5	28.6	1312	7	US-08-722-570-14
13	304	27.0	1277	7	US-08-722-570-15
14	238.5	21.2	1550	9	US-10-004-717-43
15	238.5	21.2	1957	9	US-10-004-717-8
16	207.5	18.4	993	9	US-10-004-717-47
17	207.5	18.4	3261	9	US-10-004-717-12
18	207.5	18.4	3541	9	US-10-004-717-32
19	205.5	18.2	1056	9	US-10-004-717-10
20	205.5	18.2	1393	9	US-10-004-717-45
21	198.5	17.6	675	9	US-10-004-717-15
22	197.5	17.5	485	9	US-10-004-717-59
23	192.5	17.1	1065	9	US-10-004-717-1
24	192.5	17.1	1572	9	US-10-004-717-57
25	190	16.9	501	9	US-10-004-717-13
26	187	16.6	948	9	US-10-004-717-41
27	187	16.6	2993	9	US-10-125-237-56
28	183.5	16.3	907	9	US-10-004-717-65
29	183.5	16.3	17290	9	US-09-999-121-7
30	183.5	16.3	25760	9	US-09-999-121-13
31	177.5	15.7	938	9	US-10-004-717-39
32	164	14.6	485	10	US-09-833-381-66
33	163.5	14.5	450	9	US-10-004-717-22
34	162.5	14.4	446	10	US-09-833-381-505
35	161.5	14.3	1768	10	US-09-833-381-523
36	157	13.9	804	10	US-09-749-728B-62
37	157	13.9	249487	9	US-10-026-188-3
38	153	13.6	714	10	US-09-833-381-1316
39	153	13.6	1457	9	US-09-954-531-982
40	151	13.4	2466	12	US-10-04-090-251
41	149	13.2	1042	10	US-09-833-381-1305
42	149	13.2	2264	9	US-10-004-717-49
43	148	13.1	515	9	US-10-004-717-35
44	147	13.0	447	10	US-09-864-761-27737
45	144	12.8	1476	9	US-10-004-717-16

ALIGNMENTS

RESULT 1
US-09-817-360-1
; Sequence 1, Application US/09817360
; Patent No. US20020015696A1
; GENERAL INFORMATION:
; APPLICANT: German, Michael S.
; APPLICANT: Lin, Joseph
; TITLE OF INVENTION: PRODUCTION OF PANCREATIC ISLET CELLS
; TITLE OF INVENTION: AND DELIVERY OF INSULIN
; FILE REFERENCE: UCSF-129CIP
; CURRENT APPLICATION NUMBER: US/09/817,360
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: 09/535,145
; PRIOR FILING DATE: 2000-03-24
; PRIOR APPLICATION NUMBER: 60/128,180
; PRIOR FILING DATE: 1999-04-06
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 5340
; TYPE: DNA
; ORGANISM: Homo Sapiens
US-09-817-360-1

Alignment Scores: 8.76e-97 Length: 5340
Pred. No.: 1105.00 Matches: 211
Score: 99.07%
Percent Similarity: 99.07%
Conservative: 1

Best Local Similarity: 98.60% Mismatches: 2
 Query Match: 98.05% Indels: 0
 DB: 10 Gaps: 0

US-09-595-947C-10 (1-214) x US-09-817-360-1 (1-5340)

QY 1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
 DB 3022 ATGACGCTCAACCTCGGGTGGCCACTGCTCCAAAGTACCCGCTGAGAGCGAGCGTCC 3081
 QY 21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40
 DB 3082 TTCCCCAGAGCTCGGAAGACGAGTACCTGCCCCACCTCGCGCCCGCCAGCCCACT 3141
 QY 41 ArgThrProGlyAsnGlyAlaGluAlaGluGluGlyGlyCysArgGlyAlaProArgLys 60
 DB 3142 CGCACACGCGGGAACCTGCGCAGAGCGGAGGAGGAGGCTGCGGAGGGGCCCGGAGGAAG 3201
 QY 61 LeuArgAlaArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80
 DB 3202 CTCGGGGACGCGCGGGGAGCGACCGCGGCTTAAGAGGAGTGGTGCCTGAGCAAGCAG 3261
 QY 81 ArgArgSerArgLysLysAlaAsnAspArgGluArgAsnArgMethHisAspLeuAsn 100
 DB 3262 CGACGGAGTCGGGAAAGAACGACGACGCGGAGCGCAATCGAATGCACAACTCAAC 3321
 QY 101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAlaLysLeuThr 120
 DB 3322 TCGGCACTGGAGCGCCCTGGCGGTGTCTGCCACCTTCCAGACGACGCGAAGCTCAC 3381
 QY 121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleThrAlaLeuThrGlnThrLeuArg 140
 DB 3382 AGATCGACAGCGTGGCGTTCGCCCACTACATCTGGCGGTGACTCAAGCGTGGCG 3441
 QY 141 IleAlaAspHisSerLeuTyrAlaLeuGluProProAlaProHisCysGlyGluLeuGly 160
 DB 3442 ATAGCGGACACAGCTGTACGCGTGGAGCGCGCGCGCACTGCGGGAGTGGCG 3501
 QY 161 SerProGlyProProGlyAspTrpGlySerLeuTyrSerProValSerGlnAlaGly 180
 DB 3502 AGCCAGGCGGTTCGCCGGGACTGGGGTCCCTCTACTCCCACTCTCCAGGCTGGC 3561
 QY 181 SerLeuSerProAlaAlaSerLeuGluArgProGlyLeuLeuGlyAlaThrSerSer 200
 DB 3562 AGCTGAGTCCCGCCGCTGCTGGAGAGGACGCGCGGCTGCTGGGGCCACCTCTTC 3621
 QY 201 AlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214
 DB 3622 GCCTGCTTCCAGCCAGGCTGCTGGCTTCTCAGATTTCTG 3663

RESULT 2

US-10-004-717-4
 ; Sequence 4, Application US/10004717
 ; Publication No. US20020192665A1
 ; GENERAL INFORMATION:
 ; APPLICANT: ZOEHBI, HUDA Y.
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN
 ; TITLE OF INVENTION: ATONAL ASSOCIATED SEQUENCE FOR DEAFNESS,
 ; TITLE OF INVENTION: OSTEOARTHRITIS, AND ABNORMAL CELL PROLIFERATION
 ; FILE REFERENCE: P01899US4
 ; CURRENT APPLICATION NUMBER: US/10/004,717
 ; CURRENT FILING DATE: 2002-08-16
 ; PRIOR APPLICATION NUMBER: 09/585,645
 ; PRIOR FILING DATE: 2000-06-01
 ; PRIOR APPLICATION NUMBER: 60/176,993
 ; PRIOR FILING DATE: 2000-01-19
 ; PRIOR APPLICATION NUMBER: 60/137,060
 ; PRIOR FILING DATE: 1999-06-01
 ; NUMBER OF SEQ ID NOS: 69
 ; SOFTWARE: Patentin Ver. 2.1
 ; SEQ ID NO 4
 ; LENGTH: 645

; TYPE: DNA
 ; ORGANISM: Mus musculus
 US-10-004-717-4

Alignment Scores: 2.34e-70 Length: 645
 Pred. No.: 819.00 Matches: 163
 Score: 82.33% Conservative: 14
 Percent Similarity: 75.81% Mismatches: 36
 Best Local Similarity: 72.67% Indels: 2
 Query Match: 9 Gaps: 2
 DB: 2

US-09-595-947C-10 (1-214) x US-10-004-717-4 (1-645)

QY 1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
 DB 1 ATGGCCCTCATCCCTTGGATGGCTCACCATCCAAAGTGTCCCGACAGACACAAACCT 60
 QY 21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40
 DB 61 TTCCCGGAGCTCGGACCCAGAGTGTCTCAATTCACCCCACTAGCCCACT 120
 QY 41 ArgThrProGlyAsnGlyAlaGluGluGlyGlyCysArgGlyAlaProArgLys 60
 DB 121 CTCATACCTAGGACTGCTCCGAAGCAGAAAGTGGTGTGCTGCTGCGAGGAGCTCGAGGAAG 180
 QY 61 LeuArgAlaArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80
 DB 181 CTCGCGCGCGAGCGGAGGCGCAACAGCCCAAGAGGAGTGGCAGCTCAGCAAAACAG 240
 QY 81 ArgArgSerArgLysLysAlaAsnAspArgGluArgAsnArgMethHisAspLeuAsn 100
 DB 241 CGAAGAGCGCGGCGCAAGAGGCAATGATCGGAGCGCAATCGCATGCACAACTCAAC 300
 QY 101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAlaLysLeuThr 120
 DB 301 TCGGCGCTGGATGCGTGGCGGTCTCTGCCACCTTCCCGATGACGCAAACTTACA 360
 QY 121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleThrAlaLeuThrGlnThrLeuArg 140
 DB 361 AGATCGAGACCTGCGCTTCGCCACACTACATCTGGGCACTGACTCAGAGCTCGCG 420
 QY 141 IleAlaAspHisSerLeuTyrAlaLeuGluProProAlaProHisCysGlyGluLeuGly 160
 DB 421 ATAGCGGACACAGCTTCTATGCCCGGAGCGCCCTGTGCCC ---TGTGGAGAGCTGGG 477
 QY 161 SerPro---GlyGlyProGlyAspTrpGlySerLeuTyrSerProValSerGlnAla 179
 DB 478 AGCCCGGAGGTGGCTCCCAACGGGAGTGGGGTCTATCTACTCTCCCGAGTCTCCCAAGCG 537
 QY 180 GlySerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeuLeuGlyAlaThrSer 199
 DB 538 GGTAACTGAGCCCGCAGCGCTCATTTGGAGAAATTCCTGCTGCTGAGGTGCCAGCTCC 597
 QY 200 SerAlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214
 DB 598 CCATCTATCTGCTCCCGGAGCAGCTGGTGTCTCAGACTTCTTG 642

RESULT 3

US-10-004-717-24
 ; Sequence 24, Application US/10004717
 ; Publication No. US20020192665A1
 ; GENERAL INFORMATION:
 ; APPLICANT: ZOEHBI, HUDA Y.
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN
 ; TITLE OF INVENTION: ATONAL ASSOCIATED SEQUENCE FOR DEAFNESS,
 ; TITLE OF INVENTION: OSTEOARTHRITIS, AND ABNORMAL CELL PROLIFERATION
 ; FILE REFERENCE: P01899US4
 ; CURRENT APPLICATION NUMBER: US/10/004,717
 ; CURRENT FILING DATE: 2002-08-16
 ; PRIOR APPLICATION NUMBER: 09/585,645
 ; PRIOR FILING DATE: 2000-06-01

```

; PRIOR APPLICATION NUMBER: 60/176,993
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: 60/137,060
; PRIOR FILING DATE: 1999-06-01
; NUMBER OF SEQ ID NOS: 69
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 24
; LENGTH: 861
; TYPE: DNA
; ORGANISM: Mus musculus
; US-10-004-717-24

Alignment Scores:
Pred. No.: 3 21e-70 Length: 861
Score: 819.00 Matches: 163
Percent Similarity: 82.33% Conservative: 14
Best Local Similarity: 75.81% Mismatches: 36
Query Match: 72.67% Indels: 2
DB: 9 Gaps: 2

US-09-595-947C-10 (1-214) x US-10-004-717-24 (1-861)
QY 1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
Db 160 ATGGCGCTCATCCCTTGGATGCGCTCACCATCCAAAGTGTCCCGAGACACACAACT 219
QY 21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProSerProThr 40
Db 220 TTTCCTGGAGCCCTCGGACGACGAGAGTGTCTCAATTCACATCCACCCACCTAGCCCACT 279
QY 41 ArgThrProGlyAsnGlyAlaGluGluGluGlyGlyGlyGlyGlyGlyGlyGlyGly 60
Db 280 CTATACCTAGGACTGCTCCGAGACGAGAGTGGTGTCTGCGAGGAGCTCGAGGAAG 339
QY 61 LeuArgAlaArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80
Db 340 CTCGCGCCGAGCGGAGGCGCAACAGGCGCAAGAGCGAGTGTGGCACTCAGCAACAG 399
QY 81 ArgArgSerArgGlyLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn 100
Db 400 CGAAGAAGCGGCGCAAGAGGCGCAATGTCGGAGGCGCAATGCGATGCAACACTCAAC 459
QY 101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAlaLysLeuThr 120
Db 460 TCGCGCTGTGATGCTGCGCGGTGTCTGCCACCTTCCCGATGATGACGCAACTTACA 519
QY 121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArg 140
Db 520 AAGATCAGACCTTGCCTTCGCCACAACTACATCTGGGCACTGACTCAGACGCTGCGC 579
QY 141 IleAlaAspHisSerLeuTyrAlaLeuGluProAlaProHisCysGlyGluLeuGly 160
Db 580 ATAGCGGACCACTTCTATGCGCGGAGCGCCCTGTGCC---TGTGGAGAGCTGGGG 636
QY 161 SerPro---GlyGlyProGlyAspTrpGlySerLeuTyrSerProValSerGlnAla 179
Db 637 AGCCCGGAGGTGGCTCCACGGGAGCTGGGCTCTATCTACTCCCGAGTCTCCCAAGCG 696
QY 180 GlySerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeuLeuGlyAlaThrSer 199
Db 697 GGTAACCTGAGCCCGGAGGCTCATTTGGAGAAATTCCTTGGCTTGGCTGCGGAGTCC 756
QY 200 SerAlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214
Db 757 CCATCCCTATCTGCTCCGCGGAGCACTGGTGTCTCAGACTTCTTG 801

RESULT 4
US-09-817-360-3
; Sequence 3, Application US/09817360
; Patent No. US20020015696A1
; GENERAL INFORMATION:
; APPLICANT: German, Michael S.
; APPLICANT: Lin, Joseph

```

```

; TITLE OF INVENTION: PRODUCTION OF PANCREATIC ISLET CELLS
; TITLE OF INVENTION: AND DELIVERY OF INSULIN
; FILE REFERENCE: UCSF-129CIP
; CURRENT APPLICATION NUMBER: US/09/817,360
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: 09/535,145
; PRIOR FILING DATE: 2000-03-24
; PRIOR APPLICATION NUMBER: 60/128,180
; PRIOR FILING DATE: 1999-04-06
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1861
; TYPE: DNA
; ORGANISM: Mus musculus
; US-09-817-360-3

Alignment Scores:
Pred. No.: 7 51e-70 Length: 1861
Score: 819.00 Matches: 163
Percent Similarity: 82.33% Conservative: 14
Best Local Similarity: 75.81% Mismatches: 36
Query Match: 72.67% Indels: 2
DB: 10 Gaps: 2

US-09-595-947C-10 (1-214) x US-09-817-360-3 (1-1861)
QY 1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
Db 1093 ATGGCGCTCATCCCTTGGATGCGCTCACCATCCAAAGTGTCCCGAGACACACAACT 1152
QY 21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProSerProThr 40
Db 1153 TTTCCTGGAGCCCTCGGACGACGAGAGTGTCTCAATTCACATCCACCCACCTAGCCCACT 1212
QY 41 ArgThrProGlyAsnGlyAlaGluGluGluGlyGlyGlyGlyGlyGlyGlyGlyGly 60
Db 1213 CTCATACCTAGGACTGCTCCGAGACGAGAGTGGTGTCTGCGAGGAGCTCGAGGAAG 1272
QY 61 LeuArgAlaArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80
Db 1273 CTCGCGCCGAGCGGAGGCGCAACAGGCGCAAGAGCGAGTGTGGCACTCAGCAACAG 1332
QY 81 ArgArgSerArgGlyLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn 100
Db 1333 CGAAGAAGCGGCGCAAGAGGCGCAATGTCGGAGGCGCAATGCGATGCAACACTCAAC 1392
QY 101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAlaLysLeuThr 120
Db 1393 TCGCGCTGTGATGCTGCGCGGTGTCTGCCACCTTCCCGATGATGACGCAACTTACA 1452
QY 121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArg 140
Db 1453 AAGATCAGACCTTGCCTTCGCCACAACTACATCTGGGCACTGACTCAGACGCTGCGC 1512
QY 141 IleAlaAspHisSerLeuTyrAlaLeuGluProAlaProHisCysGlyGluLeuGly 160
Db 1513 ATAGCGGACCACTTCTATGCGCGGAGCGCCCTGTGCC---TGTGGAGAGCTGGGG 1569
QY 161 SerPro---GlyGlyProGlyAspTrpGlySerLeuTyrSerProValSerGlnAla 179
Db 1570 AGCCCGGAGGTGGCTCCACGGGAGCTGGGCTCTATCTACTCCCGAGTCTCCCAAGCG 1629
QY 180 GlySerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeuLeuGlyAlaThrSer 199
Db 1630 GGTAACCTGAGCCCGGAGGCTCATTTGGAGAAATTCCTTGGCTTGGCTGCGGAGTCC 1689
QY 200 SerAlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214
Db 1690 CCATCCCTATCTGCTCCGCGGAGCACTGGTGTCTCAGACTTCTTG 1734

RESULT 5
US-10-004-717-20

```

```
; Sequence 20, Application US/10004717
; Publication No. US20020192665A1
; GENERAL INFORMATION:
; APPLICANT: ZOGHBI, HUDA Y.
; APPLICANT: YANG, QI
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN
; TITLE OF INVENTION: ATONAL ASSOCIATED SEQUENCE FOR DEAFNESS,
; TITLE OF INVENTION: OSTEOARTHRITIS, AND ABNORMAL CELL PROLIFERATION
; FILE REFERENCE: P01899US4
; CURRENT APPLICATION NUMBER: US/10/004,717
; CURRENT FILING DATE: 2002-08-16
; PRIOR APPLICATION NUMBER: 09/585,645
; PRIOR FILING DATE: 2000-06-01
; PRIOR APPLICATION NUMBER: 60/176,993
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: 60/137,060
; PRIOR FILING DATE: 1999-06-01
; NUMBER OF SEQ ID NOS: 69
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 20
; LENGTH: 790
; TYPE: DNA
; ORGANISM: chicken
US-10-004-717-20

Alignment Scores:
Pred. No.: 5,82e-29 Length: 790
Score: 388.50 Matches: 104
Percent Similarity: 54.31% Conservative: 22
Best Local Similarity: 44.83% Mismatches: 63
Query Match: 34.47% Indels: 43
DB: 9 Gaps: 10

US-09-595-947C-10 (1-214) x US-10-004-717-20 (1-790)

QY 3 ProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSerPhePro 22
DB 47 CCAGGCCGCCCGACGGCCGCGAGGCC---GTGACCAGCGCGCGCGCCCTCTCCGC 103
QY 23 Arg-----AlaSerGluAspGluVal 29
DB 104 CGCCCCCGCAGGATCGCGTGAAGCGGAGAGCGCGCGCGCGCGCGGAGGACGAACGTG 163
QY 30 ThrCysProThrSerAlaProProSerProThrArgThr---ProGlyAsnCysAlaGlu 48
DB 164 CTGCTGCTCGCCTCGCCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 223
QY 49 AlaGluGluGlyGlyCysArgGlyAlaProArgLysLysLysLysLysLysLysLys 66
DB 224 GAGGACGAGGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 283
QY 67 GlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln-----Arg 81
DB 284 GGGCGCAGCGAGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 343
QY 82 ArgSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsnSer 101
DB 344 CGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 403
QY 102 AlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAlaLysLysLysLys 121
DB 404 GGCCTGGACGCGCTGGCGGACGCTGCTGCGCCACCTTCCCGGAGGACCGCAAGCTCACA 463
QY 122 IleGluThrLeuArgPheAlaHisAsnTyrlleTrpAlaLeuThrGlnThrLeuArgIle 141
DB 464 ATCGACGCGCTGGCTTCCCGCACACTACATCTGGGCGCTCACCAGGAGCGTGGCGCTG 523
QY 142 AlaAspHisSer---LeuTyrlleAlaLeuGluProProAlaProHisCysGlyGluLeuGly 160
DB 524 GCGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 583
QY 161 SerProGlyProProGlyAspTTPGlySerLeuTyrlleSerProValSerGlnAlaGly 180
DB 161 SerProGlyProProGlyAspTTPGlySerLeuTyrlleSerProValSerGlnAlaGly 180

Db 584 AGCCCC---TCGCCCGCCTCGTCGTGG-----AGCGGC 613
QY 181 SerLeuSerProAlaAlaSerLeuGluArgProGlyLeuLeuGlyAlaThrSerSer 200
DB 614 GCGCGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 646
QY 201 AlaCys---LeuSerProGlySerLeuAlaPheSer 211
DB 647 GCCTGCACTTTATCGCCCGGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 682

RESULT 6
US-10-004-717-18
; Sequence 18, Application US/10004717
; Publication No. US20020192665A1
; GENERAL INFORMATION:
; APPLICANT: ZOGHBI, HUDA Y.
; APPLICANT: YANG, QI
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN
; TITLE OF INVENTION: ATONAL ASSOCIATED SEQUENCE FOR DEAFNESS,
; TITLE OF INVENTION: OSTEOARTHRITIS, AND ABNORMAL CELL PROLIFERATION
; FILE REFERENCE: P01899US4
; CURRENT APPLICATION NUMBER: US/10/004,717
; CURRENT FILING DATE: 2002-08-16
; PRIOR APPLICATION NUMBER: 09/585,645
; PRIOR FILING DATE: 2000-06-01
; PRIOR APPLICATION NUMBER: 60/176,993
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: 60/137,060
; PRIOR FILING DATE: 1999-06-01
; NUMBER OF SEQ ID NOS: 69
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 18
; LENGTH: 1074
; TYPE: DNA
; ORGANISM: chicken
US-10-004-717-18

Alignment Scores:
Pred. No.: 1,58e-28 Length: 1074
Score: 385.50 Matches: 104
Percent Similarity: 51.29% Conservative: 15
Best Local Similarity: 44.83% Mismatches: 60
Query Match: 34.21% Indels: 53
DB: 9 Gaps: 10

US-09-595-947C-10 (1-214) x US-10-004-717-18 (1-1074)

QY 21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProSerPro--- 39
DB 228 TTTCCCTCCCTTTCCCTCC-----CCTCAGCACCTTCCCGCGCTCCCGCTCC 275
QY 40 -----ThrArgThrProGlyAsnCysAlaGluAla 49
DB 276 GCGCGCCACTGACGGCTCGCGGTTCTCTCCCGCAGGATCCCGCGGAGCGCGCGAGCAGC 335
QY 50 GluGluGlyGlyCysArg-----GlyAlaProArgLysLeuAlaArgArgGly 66
DB 336 -----GGCGGGGTTTCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 386
QY 67 GlyArgSerArgProLysSerGluLeuAlaLeuSerLysGlnArgSerArgArgLys 86
DB 387 GGCCTGGCGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 446
QY 87 LysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsnSerAlaLeuAspAlaLeu 106
DB 447 AAAGCAACGACGCGGAGCGGAACCGCATGCACCTCAACCGCGCGCGCGCGCGCG 506
QY 107 ArgGlyValLeuProThrPheProAspAlaLysLeuThrLysIleGluThrLeuArg 126
DB 507 CGCAGGCTCTCGCGACCTTCCCGGAGCAGCACCACCACTACCAAAATGAAACCCCTGCG 566
QY 127 PheAlaHisAsnTyrlleTrpAlaLeuThrGlnThrLeuArgIleAlaAspHisSerLeu 146
DB 127 PheAlaHisAsnTyrlleTrpAlaLeuThrGlnThrLeuArgIleAlaAspHisSerLeu 146
```


QY	204	SerPro 205	Length: 1527
Db	846	TCACCT 851	Matches: 99
RESULT 10			
US-08-722-570-12	; Sequence 12, Application US/08722570		
	; Publication No. US20030044887A1		
	; GENERAL INFORMATION:		
	; APPLICANT: Anderson, David J.		
	; APPLICANT: Ma, Qifu		
	; TITLE OF INVENTION: NEUROGENIN		
	; NUMBER OF SEQUENCES: 20		
	; ADDRESS: Flehr, Hobbach, Test, Albritton & Herbert		
	; STREET: Four Embarcadero Center, Suite 3400		
	; CITY: San Francisco		
	; STATE: California		
	; COUNTRY: United States		
	; ZIP: 94111-4187		
	; COMPUTER READABLE FORM:		
	; MEDIUM TYPE: Floppy disk		
	; OPERATING SYSTEM: PC-DOS/MS-DOS		
	; SOFTWARE: Patent In Release #1.0, Version #1.30		
	; CURRENT APPLICATION DATA:		
	; FILING DATE: 27-SEP-1996		
	; CLASSIFICATION: 5365		
	; ATTORNEY/AGENT INFORMATION:		
	; NAME: Silva, Robin M.		
	; REGISTRATION NUMBER: 38,304		
	; REFERENCE/DOCKET NUMBER: A-63902/REF/RMS		
	; TELECOMMUNICATION INFORMATION:		
	; TELEPHONE: (415) 781-1989		
	; TELEFAX: (415) 398-3249		
	; INFORMATION FOR SEQ ID NO: 12:		
	; SEQUENCE CHARACTERISTICS:		
	; LENGTH: 1527 base pairs		
	; TYPE: nucleic acid		
	; STRANDEDNESS: unknown		
	; TOPOLOGY: unknown		
	; MOLECULE TYPE: DNA		
US-08-722-570-12	; Sequence 13, Application US/08722570		
	; Publication No. US20030044887A1		
	; GENERAL INFORMATION:		
	; APPLICANT: Anderson, David J.		
	; APPLICANT: Ma, Qifu		
	; TITLE OF INVENTION: NEUROGENIN		
	; NUMBER OF SEQUENCES: 20		
	; ADDRESS: Flehr, Hobbach, Test, Albritton & Herbert		
	; STREET: Four Embarcadero Center, Suite 3400		
	; CITY: San Francisco		
	; STATE: California		
	; COUNTRY: United States		
	; ZIP: 94111-4187		
	; COMPUTER READABLE FORM:		
	; MEDIUM TYPE: Floppy disk		
	; OPERATING SYSTEM: PC-DOS/MS-DOS		
	; SOFTWARE: Patent In Release #1.0, Version #1.30		
	; CURRENT APPLICATION DATA:		
	; FILING DATE: 27-SEP-1996		
	; CLASSIFICATION: 5365		
	; ATTORNEY/AGENT INFORMATION:		
	; NAME: Silva, Robin M.		
	; REGISTRATION NUMBER: 38,304		
	; REFERENCE/DOCKET NUMBER: A-63902/REF/RMS		
	; TELECOMMUNICATION INFORMATION:		
	; TELEPHONE: (415) 781-1989		
	; TELEFAX: (415) 398-3249		
	; INFORMATION FOR SEQ ID NO: 13:		
	; SEQUENCE CHARACTERISTICS:		
	; LENGTH: 1527 base pairs		
	; TYPE: nucleic acid		
	; STRANDEDNESS: unknown		
	; TOPOLOGY: unknown		
	; MOLECULE TYPE: DNA		
US-08-722-570-12	; Sequence 14, Application US/08722570		
	; Publication No. US20030044887A1		
	; GENERAL INFORMATION:		
	; APPLICANT: Anderson, David J.		
	; APPLICANT: Ma, Qifu		
	; TITLE OF INVENTION: NEUROGENIN		
	; NUMBER OF SEQUENCES: 20		
	; ADDRESS: Flehr, Hobbach, Test, Albritton & Herbert		
	; STREET: Four Embarcadero Center, Suite 3400		
	; CITY: San Francisco		
	; STATE: California		
	; COUNTRY: United States		
	; ZIP: 94111-4187		
	; COMPUTER READABLE FORM:		
	; MEDIUM TYPE: Floppy disk		
	; OPERATING SYSTEM: PC-DOS/MS-DOS		
	; SOFTWARE: Patent In Release #1.0, Version #1.30		
	; CURRENT APPLICATION DATA:		
	; FILING DATE: 27-SEP-1996		
	; CLASSIFICATION: 5365		
	; ATTORNEY/AGENT INFORMATION:		
	; NAME: Silva, Robin M.		
	; REGISTRATION NUMBER: 38,304		
	; REFERENCE/DOCKET NUMBER: A-63902/REF/RMS		
	; TELECOMMUNICATION INFORMATION:		
	; TELEPHONE: (415) 781-1989		
	; TELEFAX: (415) 398-3249		
	; INFORMATION FOR SEQ ID NO: 14:		
	; SEQUENCE CHARACTERISTICS:		
	; LENGTH: 1527 base pairs		
	; TYPE: nucleic acid		
	; STRANDEDNESS: unknown		
	; TOPOLOGY: unknown		
	; MOLECULE TYPE: DNA		
US-08-722-570-12	; Sequence 15, Application US/08722570		
	; Publication No. US20030044887A1		
	; GENERAL INFORMATION:		
	; APPLICANT: Anderson, David J.		
	; APPLICANT: Ma, Qifu		
	; TITLE OF INVENTION: NEUROGENIN		
	; NUMBER OF SEQUENCES: 20		
	; ADDRESS: Flehr, Hobbach, Test, Albritton & Herbert		
	; STREET: Four Embarcadero Center, Suite 3400		
	; CITY: San Francisco		
	; STATE: California		
	; COUNTRY: United States		
	; ZIP: 94111-4187		
	; COMPUTER READABLE FORM:		
	; MEDIUM TYPE: Floppy disk		
	; OPERATING SYSTEM: PC-DOS/MS-DOS		
	; SOFTWARE: Patent In Release #1.0, Version #1.30		
	; CURRENT APPLICATION DATA:		
	; FILING DATE: 27-SEP-1996		
	; CLASSIFICATION: 5365		
	; ATTORNEY/AGENT INFORMATION:		
	; NAME: Silva, Robin M.		
	; REGISTRATION NUMBER: 38,304		
	; REFERENCE/DOCKET NUMBER: A-63902/REF/RMS		
	; TELECOMMUNICATION INFORMATION:		
	; TELEPHONE: (415) 781-1989		
	; TELEFAX: (415) 398-3249		
	; INFORMATION FOR SEQ ID NO: 15:		
	; SEQUENCE CHARACTERISTICS:		
	; LENGTH: 1527 base pairs		
	; TYPE: nucleic acid		
	; STRANDEDNESS: unknown		
	; TOPOLOGY: unknown		
	; MOLECULE TYPE: DNA		
US-08-722-570-12	; Sequence 16, Application US/08722570		
	; Publication No. US		

RESULT 13

US-08-722-570-15
 ; Sequence 15, Application US/08722570
 ; Publication No. US20030044887A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Anderson, David J.
 ; APPLICANT: Ma, Olufo
 ; TITLE OF INVENTION: NEUROGENIN
 ; NUMBER OF SEQUENCES: 20
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
 ; STREET: Four Embarcadero Center, Suite 3400
 ; CITY: San Francisco
 ; STATE: California
 ; COUNTRY: United States
 ; ZIP: 94111-4187
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/722,570
 ; FILING DATE: 27-SEP-1996
 ; CLASSIFICATION: 5365
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Silva, Robin M.
 ; REGISTRATION NUMBER: 38,304
 ; REFERENCE/DOCKET NUMBER: A-63902/RFT/RMS
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (415) 781-1989
 ; TELEFAX: (415) 398-3249
 ; TELEX: 910 277299
 ; INFORMATION FOR SEQ ID NO: 15:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1277 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: unknown
 ; TOPOLOGY: unknown
 ; MOLECULE TYPE: DNA
 ; US-08-722-570-15

Alignment Scores:
 Pred. No.: 1,266-20 Length: 1277
 Score: 304.00 Matches: 86
 Percent Similarity: 57.29% Conservative: 24
 Best Local Similarity: 44.79% Mismatches: 64
 Query Match: 26.97% Indels: 20
 DB: 7 Gaps: 8

US-09-595-947C-10 (1-214) x US-08-722-570-15 (1-1277)

QY 24 AlaSerGluAspGluValThrCysProThrSerAlaProProSerProThrArgThrPro 43
 DB 298 AGCTCGGACGATGAG-----CAGCTACACAGTCCGACGACCGCGCGTCAAG--- 345
 QY 44 GlyAsnCysAlaGluAlaGluGlyCysArgGlyAlaProArgLysLeuArgAla 63
 DB 346 -----CACCTGACGAGGACCGGACCGGAGGAGGAGAACACCGCGCGATGC 393
 QY 64 ArgArgGlyCysArgSerArgProLysSerGluLeuAlaLeuSerLysGlnArgArgSer 83
 DB 394 AGG-----AGGAGCGGACCGCGGAGACACCGTCTG---AAGATCAAGAGACG 441
 QY 84 ArgArgLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsnSerAlaLeu 103
 DB 442 CGCGCGTTAAAGCAATAACCGGAGAGAGAAATCGCATCCACCACTGAATATGGCTC 501
 QY 104 AspAlaLeuGlyValLeuProThrPheProAspAlaLysLeuThrLysIleGlu 123
 DB 502 GATTCTCTGAGGAGGTTCTACCGTCATTACCGGAGAGCGCAACCACTCAAGATAGAG 561

QY 124 ThrLeuArgPheAlaHisAsnTyrlleTAlaLeuThrGlnThrLeuArgIleAlaAsp 143
 DB 562 ACCTTGGCGCTTGGCCACCACTACATCTGGGCTCTTAGGAAACTTTGGCCCTGGCGAC 621
 QY 144 -----HisSerLeuTyrlAlaLeuGluProProAlaProHisCysGlyGlu-----Leu 159
 DB 622 CAGCTGCACGG-ATCTAC---TTCCACCCCGCAGCAGCAGCATATGGTACAGGACTCCTA 677
 QY 160 GlySerProGlyGlyProProGlyAspTrpGlySerLeuTyrlSerProValSerGlnAla 179
 DB 678 TCCTTCCCTGA-GCCCTCTCTGGAGCTGCAGCTCGCTCCCACTCTTCGGACT 736
 QY 180 GlySerLeuSerProAlaAlaSerLeuGluArgProGlyLeu---LeuGlyAlaThr 198
 DB 737 CCTTCTCCCGCAGCAGCCCTGCCAGCTCCACCTCGACAGTATTGAGTACTGCGACGCC 796
 QY 199 SerSerAlaCysLeuSerProGlySerLeuAlaPhe 210
 DB 797 CTGAGCTCGGCTTGAACCCCTTCATGCTGCGCCTTT 832

RESULT 14

US-10-004-717-43
 ; Sequence 43, Application US/10004717
 ; Publication No. US20020192665A1
 ; GENERAL INFORMATION:
 ; APPLICANT: ZOGHEI, HUDA Y.
 ; APPLICANT: YANG, QI
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN
 ; TITLE OF INVENTION: ATONAL ASSOCIATED SEQUENCE FOR DEAFNESS,
 ; TITLE OF INVENTION: OSTEOARTHRITIS, AND ABNORMAL CELL PROLIFERATION
 ; FILE REFERENCE: P01899US4
 ; CURRENT APPLICATION NUMBER: US/10/004,717
 ; CURRENT FILING DATE: 2002-08-16
 ; PRIOR APPLICATION NUMBER: 09/585,645
 ; PRIOR FILING DATE: 2000-06-01
 ; PRIOR APPLICATION NUMBER: 60/176,993
 ; PRIOR FILING DATE: 2000-01-19
 ; PRIOR APPLICATION NUMBER: 60/137,060
 ; PRIOR FILING DATE: 1999-06-01
 ; NUMBER OF SEQ ID NOS: 69
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 43
 ; LENGTH: 1550
 ; TYPE: DNA
 ; ORGANISM: Mus musculus
 ; US-10-004-717-43

Alignment Scores:
 Pred. No.: 3e-14 Length: 1550
 Score: 238.50 Matches: 78
 Percent Similarity: 39.29% Conservative: 21
 Best Local Similarity: 30.95% Mismatches: 72
 Query Match: 21.16% Indels: 81
 DB: 9 Gaps: 8

US-09-595-947C-10 (1-214) x US-10-004-717-43 (1-1550)

QY 19 ArgSerPheProArgAlaSerGluAsp-----GluValThrCysProThrSerAlaPro 36
 DB 106 AGAAAGTTTGCTAGACAATGTGAGGACCAACAAATAAGAAACACGAGAGCTTTCCA 165
 QY 37 -----ProSerProThrArgThrProGlyAsnCysAlaGluAla 49
 DB 166 AAACAAGTTGCTCTCGAGGAAGAGCATTAAGAGGCGCCCTGGAGAGAAACCGAGAAA 225
 QY 50 Glu-----GluGlyCysArgGlyAlaProArg 59
 DB 226 GAAGAGGAGGAAGACAGACAGAGAGGAAGAAGATGACAATGGC----- 267
 QY 60 LysLeuArgAlaArgArgGlyArgSerArgProLysSerGluLeuAlaLeuSerLys 79
 DB 268 ---TTGTCAGAGAGGGGGGCTCAGGAAAAAAGAACCAACCACTAGCTGGAAGG 324

QY 80 GlnArgSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeu 99
 Db 325 GTC---AAGTTCAGGAGAGGAGGCTAATGCGCGGAGAGGAAACCGGATGCACGGCCTC 381
 QY 100 AsnSerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAlaLysLeu 119
 Db 382 AATGATGCTCTGCACAAATTTGCGAAAGTGTCTCCCTGTTACTCTAAACCCCAAAACTG 441
 QY 120 ThrLysIleGluThrLeuArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeu 139
 Db 442 TCCAAATAGAACTTACGACTGGCCAAATATACATCTGGGCACCTTCTGAAATCTG 501
 QY 140 ArgIle-----AlaAspHis-----SerLeuTyrAlaLeuGluProAla 153
 Db 502 AGGATTGGCAAGAGACCGGATCTGCTCACGTTGTCCTCAAAACTATGCAAAAGTCTTTCC 561
 QY 141 -----AlaAspHis-----SerLeuTyrAlaLeuGluProAla 153
 Db 562 CAGCAACTACAACTGTGTGGCAGGCTGCTTACAGCTCAACGCCAGAGTTCCTGATG 621
 QY 142 -----AlaAspHis-----SerLeuTyrAlaLeuGluProAla 153
 Db 622 GGTCAAGGTGGGAGGCTGCCACCAACAGGTCAACCTACTCCACATTCACCCACCC 681
 QY 154 ProHisCysGlyGluLeuGlySerProGlyGlyProGlyProGlyAspTrpGlySerLeuTyr 173
 Db 682 TACCACAGCCCTGAGTGGCCACTCCGCCAGG-----CAT 717
 QY 174 SerProValSerGlnAlaGlySerLeuSerProAlaAlaSerLeuGluArgProGly 193
 Db 718 GGGACTCTTGATATTCGAATCCAAAGTCCATGAACCCCTACAACTACTGCAGTGCATATGAATCC 777
 QY 194 LeuLeuGlyAlaThrSerSerAlaCysLeuSerPro 205
 Db 778 TTCTATGAAGTACCTCCCTGAGTGTGCCAGCCCT 813

RESULT 15

US-10-004-717-8
 ; Sequence 8, Application US/10004717
 ; Publication No. US20020192665A1
 ; GENERAL INFORMATION:
 ; APPLICANT: ZOGHEI, HUDA Y.
 ; APPLICANT: YANG, QI
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN
 ; TITLE OF INVENTION: ATONAL ASSOCIATED SEQUENCE FOR DEAFNESS,
 ; FILE OF INVENTION: OSTEOARTHRITIS, AND ABNORMAL CELL PROLIFERATION
 ; FILE REFERENCE: P01899US4
 ; CURRENT APPLICATION NUMBER: US/10/004,717
 ; CURRENT FILING DATE: 2002-08-16
 ; PRIOR APPLICATION NUMBER: 09/585,645
 ; PRIOR FILING DATE: 2000-06-01
 ; PRIOR APPLICATION NUMBER: 60/176,993
 ; PRIOR FILING DATE: 2000-01-19
 ; PRIOR APPLICATION NUMBER: 60/137,060
 ; PRIOR FILING DATE: 1999-06-01
 ; NUMBER OF SEQ ID NOS: 69
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 8
 ; LENGTH: 1957
 ; TYPE: DNA
 ; ORGANISM: Mus musculus
 US-10-004-717-8

Alignment Scores:

Pred. No.:	3.88e-14	Length:	1957
Score:	238.50	Matches:	78
Percent Similarity:	39.29%	Conservative:	21
Best Local Similarity:	30.95%	Mismatches:	72
Query Match:	21.16%	Indels:	81
DB:	9	Gaps:	8

QY 19 ArgSerPheProArgAlaSerGluAsp-----GluValThrCysProThrSerAlaPro 36
 Db 182 AGAAAGTTTCTAGACAATGTGAGGACCAGAAACAATAAAGAAACCCAGAGAGCTTTTCCA 241
 QY 37 -----ProSerProThrArgThrProGlyAsnCysAlaGluAla 49
 Db 242 AAACAAGTTTCTCTCTGAGGAAAGAGCATTAAGAGGGCCCTGGAGAGAAACCGAGAAA 301
 QY 50 Glu-----GluGlyGlyCysArgGlyAlaProArg 59
 Db 302 GAAGAGGAGGAAGAAGACAGAGAGGAAGAAGATGAGATGGC----- 343
 QY 60 LysLeuArgAlaArgArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLys 79
 Db 344 ---TTGTCCAGAGAGGGGGCTCAGGAAAAAAGAACACCAAACTACGACTGGAAGG 400
 QY 80 GlnArgSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeu 99
 Db 401 GTC---AAGTTCAGGAGACAGGAAGCTAATGCGCGGAGAGGAACCGGATGCACGGCCTC 457
 QY 100 AsnSerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAlaLysLeu 119
 Db 458 AATGATGCTCTGGCAATTTGGCAAAAGTGTCTCCCTGTTACTCTAAACCCCAAAACTG 517
 QY 120 ThrLysIleGluThrLeuArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeu 139
 Db 518 TCCAAAATAGAACTTTACGACTGGCCAAAAATATACATCTGGGCACCTTCTGAAATCTG 577
 QY 140 ArgIle----- 141
 Db 578 AGGATTGGCAAGAGACCGGATCTGCTCACGTTGTCCTCAAAACTATGCAAAAGTCTTTCC 637
 QY 141 -----AlaAspHis-----SerLeuTyrAlaLeuGluProAla 153
 Db 638 CAGCCAACTACAACTGTGTGGCAGGCTGCTTACAGCTCAACGCCAGAGTTCCTGATG 697
 QY 142 -----AlaAspHis-----SerLeuTyrAlaLeuGluProAla 153
 Db 698 GGTCAAGGTGGGAGGCTGCCACCAACAGGTCAACCTACTCCACATTCACCCACCC 757
 QY 154 ProHisCysGlyGluLeuGlySerProGlyGlyProGlyProGlyAspTrpGlySerLeuTyr 173
 Db 758 TACCACAGCCCTGAGTGGCCACTCCCCAGG-----CAT 793
 QY 174 SerProValSerGlnAlaGlySerLeuSerProAlaAlaSerLeuGluArgProGly 193
 Db 794 GGCACCTTTGATAATTCCAAGTCCATGAACCCCTACAACTACTGCAGTGCATATGAATCC 853
 QY 194 LeuLeuGlyAlaThrSerSerAlaCysLeuSerPro 205
 Db 854 TTCTATGAAGTACCTCCCTGAGTGTGCCAGCCCT 889

Search completed: April 9, 2003, 09:58:37
 Job time : 586 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 8, 2003, 22:56:59 ; Search time 1739 Seconds
(without alignments)
1993.004 Million cell updates/sec

Title: US-09-595-947C-10
Perfect score: 1127
Sequence: 1 MTQPSGAPVQVITRETERS.....LGATSSACLSFGSLAFSDFL 214

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xih
-Q/cgn2.1/USPTO.spool/US09595947/runat_07042003_160405_20073/app_query.fasta_1.391
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=PCT -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09595947 -CGN_1_1906 -runat_07042003_160405_20073 -NCPU=6 -ICPU=3
-NO_XLPAY -NO_MMAB -LARGQUERY -NEG_SCORES=0 -WAIT -LONGLOP -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_invi:*
20: em_gss_pln:*
21: em_gss_vit:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_Other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	809	71.8	1540	11	AK008017 Mus muscu
2	676.5	60.0	593	17	AZ296526 RPCI-23-1
3	327	29.0	1098	14	BQ169355 STR00303
4	317.5	28.2	632	13	BJ010277 BJ010277
5	310.5	27.6	600	12	BG808248 2083-52 M
6	286.5	26.3	1039	12	BE780690 601469349
7	285.5	25.3	1037	17	CNS03VJ9 Tetraodon
8	280	24.8	629	13	BQ093114 BJ093114
9	278	24.7	595	13	BJ030202 BJ030202
10	268.5	23.8	804	9	AL540071 AL540071
11	267.5	23.7	401	10	AW147434 da02h12.y
12	251	22.3	588	10	AV673464 AV673464
13	249	22.1	490	10	AV995230 AV995230
14	243	21.6	814	14	BQ178789 UI-M-EVO-
15	239	21.2	1003	17	CNS02IB1 Tetraodon
16	236.5	20.1	742	14	BQ572268 UI-M-FD0-
17	226.5	20.1	770	14	BQ572426 UI-M-FD0-
18	224.5	19.9	688	12	BG699059 602678696
19	224.5	19.9	901	14	BQ424098 AGENCOURT
20	224.5	19.9	1967	11	BC022560 Homo sap1
21	222.5	19.7	953	9	AU067624 AU067624
22	219.5	19.5	724	12	BE783567 601471617
23	219	19.4	875	17	CNS02BRP Tetraodon
24	217.5	19.3	990	14	BQ086669 AGENCOURT
25	217.5	19.3	1065	12	BE777799 601463047
26	215	19.1	752	13	BI732893 603355021
27	215	19.1	809	12	BE783082 601470669
28	215	19.1	1009	13	BM559083 AGENCOURT
29	214	19.0	994	10	BB610830 BB610830
30	213.5	18.9	1055	9	AL535259 AL535259
31	212	18.8	600	12	BG802832 0191-69 M
32	212	18.8	651	12	BG296688 602393635
33	212	18.8	657	14	BM728742 UI-E-E01-
34	212	18.8	678	12	BG296579 602394490
35	212	18.8	692	13	BI737697 603358625
36	212	18.8	782	13	BI732663 603354913
37	212	18.8	835	12	BG298117 602396478
38	212	18.8	951	14	BQ891444 AGENCOURT
39	212	18.8	968	13	BI731880 603353625
40	212	18.8	1007	10	BB610207 BB610207
41	212	18.8	1017	14	BQ895731 AGENCOURT
42	212	18.8	1121	12	BG294629 602391973
43	212	18.8	2494	11	AK018781 Mus muscu
44	212	18.8	2495	11	AK005073 Mus muscu
45	209.5	18.6	879	13	BM450929 AGENCOURT

ALIGNMENTS

RESULT 1
AK008017

LOCUS

DEFINITION

ACCESION

VERSION

KEYWORDS

SOURCE

ORGANISM

LOCUS

DEFINITION

ACCESION

VERSION

KEYWORDS

SOURCE

ORGANISM

AK008017 1540 bp mRNA linear HTC 19-JAN-2002
Mus musculus adult male small intestine cDNA, RIKEN full-length
enriched library, clone:2010001M19:atonal homolog 5 (Drosophila),
full insert sequence.
AK008017
AK008017.1 GI:12841941
HTC; CAP trapper.
Mus musculus (strain:C57BL/6J) adult male small intestine cDNA to
mRNA, clone:lib.RIKEN full-length enriched mouse cDNA library
clone:2010001M19.
Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

Qy 81 ArgArgSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn 100
 |||||
 Db 481 CGAAGAAGCCGGCGCAGAGAGCCCAATGATCGGAGCGCAATCGCATGCAACCTCAAC 540
 |||||
 Qy 101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThr 120
 |||||
 Db 541 TCGCGCTGGATGGCTGGCGGGTCTCTCCCACTTCGCGGATGACGCCAACAATTACA 600
 |||||
 Qy 121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArg 140
 |||||
 Db 601 AAGTCGAGACCTTCGCTTCGCCACAACTACATCTGGGACATGACTCAGACGCTGGCG 660
 |||||
 Qy 141 IleAlaAspHisSerLeuThrAlaLeuGluProAlaProHisCysGlyGluLeuGly 160
 |||||
 Db 661 ATACGGACCAACAGCTGTATGGCCGGAGCCCTCTGTGCC---TGTGGAGAGCTGGGG 717
 |||||
 Qy 161 SerPro---GlyGlyProProGlyAspTrpGlySerLeuThrSerProValSerGlnAla 179
 |||||
 Db 718 AGCCCGGAGGTGCTCCACGGGAGCTGGGGCTCTATCTACTCCCACTCTCCCAAGCG 777
 |||||
 Qy 180 GlySerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeuLeuGlyAlaThrSer 199
 |||||
 Db 778 GGTAACTGACCCACGCGCTCTATGGAGGAATTCCTGGCTGCGAGTGCACGCTCC 837
 |||||
 Qy 200 SerAlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214
 |||||
 Db 838 CCATCTATCTGCTCCCGGGAGCACTGTGTCTCAGACTTCTTG 882
 |||||

RESULT 2
 AZ296526/c 593 bp DNA linear GSS 27-JUL-2000
 LOCUS RPCI-23-160G18.TV RPCI-23 Mus musculus genomic clone RPCI-23-160G18
 DEFINITION , DNA sequence.
 ACCESSION AZ296526.1 GI:9538311
 VERSION 1
 KEYWORDS house mouse.
 SOURCE Mus musculus
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 593)
 Zhao,S., Nierman,W., Feldblum,T., Malek,J., Shatsman,S., Akinret
 B., Levins,M., Mcgann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
 and Fraser,C.M.
 Mouse BAC End Sequences from Library RPCI-23
 Unpublished (1999)
 Other_GSSs: RPCI-23-160G18.TJ
 Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: szhao@tigr.org
 Clones are derived from the mouse BAC library RPCI-23. For BAC
 library availability, please contact Pieter de Jong
 (pieter@dejong.med.buffalo.edu). Clones may be purchased from
 BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>)
 or from Resea ch Genetics (info@resgen.com). BAC end page:
http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
 Plate: 160 row: 6 column: 18
 Seg primer: T7
 Class: BAC ends.
 Location/Qualifiers
 1. 593
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="RPCI-23-160G18"
 /clone_lib="RPCI-23"
 /sex="Female"
 /lab_host="DH10B"
 /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:

FEATURES
 source
 1. 593
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="RPCI-23-160G18"
 /clone_lib="RPCI-23"
 /sex="Female"
 /lab_host="DH10B"
 /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:

ECORI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
 brain genomic DNA was isolated and partially digested
 with a combination of EcoRI and EcoRI Methyase. Size
 selected DNA was cloned into the pBACe3.6 vector at the
 EcoRI sites. The ligation products were transformed into
 DH10B electrocompetent cells (BRL Life Technologies).
 BASE COUNT 88 a 159 c 213 g 133 t
 ORIGIN

Alignment Scores:
 Pred. No.: 1.51e-43 Length: 593
 Score: 676.50 Matches: 136
 Percent Similarity: 82.68% Conservative: 12
 Best Local Similarity: 75.98% Mismatches: 30
 Query Match: 60.03% Indels: 2
 DB: 17

US-09-595-947C-10 (1-214) x AZ296526 (1-593)

Qy 3 ProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSerPhePro 22
 |||||
 Db 547 CCTCATCCTTGGATGCGCTCACCATCAAGTGTCCCGAGACACAAACCTTTTCCC 488
 |||||
 Qy 23 ArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThrArgThr 42
 |||||
 Db 487 GGAGCTCGGACCAAGAGTGTCTCAATTCACCCCACTAGCCCACTCTCATA 428
 |||||
 Qy 43 ProGlyAsnCysAlaGluAlaGluGlyGlyCysArgGlyAlaProArgLysLeuArg 62
 |||||
 Db 427 CCTAGGACTGTCTCGAAGACAGAGTGGTGTACTCCGAGGACCTCAGGAAGTCCGC 368
 |||||
 Qy 53 AlaArgArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGlnArg 82
 |||||
 Db 367 GCGGACGCGGAGGCGGCAAGCCCAAGAGTGGCTGCTCCGAGGACCTCAGGAAGTCCGC 308
 |||||
 Qy 83 SerArgArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsnSerAla 102
 |||||
 Db 307 AGCCGGCGCAAGAAGCCCAATGATCGGAGCGCAATCGCATGCACAACTCAACTCGGCG 248
 |||||
 Qy 103 LeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThrLysIle 122
 |||||
 Db 247 CTGGATGGCTGCGGGGTCTCTGCCACCTTCCCGGATGAGCCAAACTTACAAAGATC 188
 |||||
 Qy 123 GluThrLeuArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArgIleAla 142
 |||||
 Db 187 GAGACCTGCGCTTCTCCCACTACATCTGGCACTGACTCAGCGCTGCGCATAGCG 128
 |||||
 Qy 143 AspHisSerLeuThrAlaLeuGluProAlaProHisCysGlyGluLeuGlySerPro 162
 |||||
 Db 127 GACCACAGCTTCTATGGCCGAGCCCTCTGTGCCCTG---GGAGAGCTGGGAGCCCC 72
 |||||
 Qy 163 GlyGlyProProGlyAspTrpGlySerLeuThrSerProValSerGlnAlaGlySer 181
 |||||
 Db 71 GGAGGGGTCCCAACGGGACTGGGCTCATCTACTCCCACTCTCCCAAGCGGAACC 15
 |||||

RESULT 3
 BQ169355 1098 bp mRNA linear EST 01-MAY-2002
 LOCUS STR00303 segmentation stage cDNA library Danio rerio cDNA clone
 DEFINITION CB260 5', similar to NEUROGENIN 1, mRNA sequence.
 ACCESSION BQ169355
 VERSION BQ169355.1 GI:20376783
 KEYWORDS EST.
 SOURCE zebrafish.
 ORGANISM Danio rerio
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes
 ; Cyprinidae; Danio.
 1 (bases 1 to 1098)
 Loppin,B., Pflumio,S., Steffan,T., Heyer,V., Furthauer,M., Thisse
 C. and Thisse,B.
 Expression of the zebrafish genome during embryogenesis (2002)
 Unpublished (2002)


```

QY 55 ArgGlyAlaProArgLysLeuArgAlaArgArgGlyGlyArgSerArgProLysSerGlu 74
Db 226 -----GGCAACAGAAAGGAGCGCGCCAGCGCGCGCGCGC-----TGCAG 270
QY 75 LeuAlaLeuSerLysArgGlnArgSerArgArgLysLysAlaAsnAspArgGluArgAsn 94
Db 271 GCGGTGCGTGCAGTGTGAAAGAAAGACCGCGCGCTGAAGCTAACGACCGGGAACGGAAC 330
QY 95 ArgMetHisAspLeuAsnSerAlaLeuAspAlaLeuArgGlyValLeuProThrPhePro 114
Db 331 CCGATGCATAACCTGAACGACCGCTGGATGAGCTGCGCGCGCTCTCGCGCCCTTCGG 390
QY 115 AspAspAlaLeuThrLysLysLeuThrLeuArgPheAlaHisAsnTyrIleThrPala 134
Db 391 GACGAACCAACCTGACCAAGATGAGACTCTGCGCTTTGCCACAACTACATCTGGGT 450
QY 135 LeuThrGlnThrLeuArgGlnAlaAspHisSerLeuTyrAla-----LeuGluPro 151
Db 451 CTGTCCGAGACCATCGCATCGACAGAC-----CTGACGGCGGGAGGGGCAACGACACC 504
QY 152 ProAla-----ProHisCys---GlyGluLeuGlySerProGlyGlyProGly 167
Db 505 CQTGCTGCTTAACCCCTGCTGCTCGCGGAGACCCGAGTCCAGCCAGTGCACCCCTGC 564
QY 168 AspTyrGlySerLeuTyrSerProValSerGlnAlaGlySerLeuSerProAlaAlaSer 187
Db 565 AGCTGG-----AGCCCCAGGGATCTCTGTCGCGCTCTCTCCCGCG----- 606
QY 188 LeuGluGluArgProGlyLeuLeuGlyAlaThrSerSerAlaCysLeuSerProGlySer 207
Db 607 -----TACTGCGCGTNCAGCCCGGCGCAGC 630

RESULT 5
BG808248 600 bp mRNA linear EST 20-DEC-2001
LOCUS 2083-52 Mouse E14.5 retina lambda ZAP II Library Mus musculus cDNA,
DEFINITION mRNA sequence.
ACCESSION BG808248
VERSION BG808248.1 GI:17955225
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Mur.X., Zhao,S., Pershad,R., Hsieh,T.-F., Scarpa,A., Wang,S.W.,
AUTHORS White,R.A., Beremand,P.D., Thomas,T.L., Gan,L. and Klein,W.H.
TITLE Gene expression in the developing mouse retina by EST sequencing
and microarray analysis
JOURNAL Nucleic Acids Res. 29 (24), 4983-4993 (2001)
MEDLINE 21671825
COMMENT Contact: Klein WH
Department of Biochemistry and Molecular Biology
University of Texas M.D. Anderson Cancer Center
Box 117, 1515 Holcombe Blvd., Houston, TX 77030, USA
Tel: 713 792 3646
Fax: 713 790 0329
FEATURES
1..600 Location/Qualifiers
source
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone_lib="Mouse E14.5 retina lambda ZAP II Library"
/tissue_type="neural retina"
/dev_stage="embryonic day 14.5 post-fertilization"
/notes="Vector: pAMP10 (Gibco); Cloned unidirectionally.
primer: Oligo dT. RNA Isolation: cytoplasmic RNA preps
(Manniatidis); Cloning Technique: CUA Cloning (Clontech,
Life Technologies); Average insert size: 1.8 Kb;
Insertion site: TACGCCACTGAATCTGATG---. Other
information regarding entire library may be found at
http://pga.swmed.edu/Data/Libraries/microarray_cdna_librar
ies.htm."
BASE COUNT 94 a 238 c 161 g 107 t

```

```

ORIGIN
Alignment Scores: 6.67e-15 Length: 600
Pred. No.: 310.50 Matches: 73
Score: 58.39% Conservative: 7
Best Local Similarity: 53.28% Mismatches: 35
Query Match: 27.55% Indels: 23
DB: 12 Gaps: 3
US-09-595-947C-10 (1-214) x BG808248 (1-600)
QY 91 ArgGluArgAsnArgMetHisAspLeuAsnSerAlaLeuAspAlaLeuArgGlyValLeu 110
Db 1 CCGGAGCGCAACCGCATGCACACCTAAACGCGCGCTGAGCGCGCTGCGGAGGTGCTG 60
QY 111 ProThrPheProAspAlaLysLeuThrLysLysLeuThrLysLysLeuThrLysLysLeu 130
Db 61 CCCACCTTCCCGGAGGATGCCAAGCTCAGAAAGATCGAGACGCTGCGCTTCGCCCAAT 120
QY 131 TyrIleThrAlaLeuThrGlnThrLeuArgGlnAlaAspHisSerLeuTyrAla----- 148
Db 121 TACATCTGGCGCTCACCAGAGACTCTGCGCTGGCGGACCATGCGCGCGCGCGGTGGC 180
QY 149 -----LeuGluProProAlaProHisCysGlyGlu 158
Db 181 CTCAGGGGGCGCTCTTCACGAGGCGGTGCTCTCTGAGCCGCGGAGCTGCGCTCGCGCC 240
QY 159 LeuGlySerProGlyGlyProGlyProGlyAspTrpGlySerLeuTyrSerProValSerGln 178
Db 241 AGCGGGGACAGCCCTTCTCCAGCTTCTCTCTGAGAGTGCACCAACAGCCCGCGCTCATCC 300
QY 179 AlaGlySerLeuSerPro-----AlaAlaSerLeuGluGluArgPro 192
Db 301 TCCAACTCCAGCTCCCACTACAGCTGCACCTTTATCCCGCGCTAGACCC-CGGGTACACGT 359
QY 193 GlyLeuGlyAlaThrSerSerAla-----CysLeuSerPro 205
Db 360 GGACTACTGGCAGCCCGCCACCTCCGGAGAGAGCATCGTTATGCGCTCACCT 410
RESULT 6
BE780690 1039 bp mRNA linear EST 20-OCT-2000
LOCUS 601469349Fl NIH_MGC_67 Homo sapiens cDNA clone IMAGE:3872172 5',
DEFINITION mRNA sequence.
ACCESSION BE780690
VERSION BE780690.1 GI:10201888
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE NIH-MGC http://mgi.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Inceye Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: LLAM9626 row: b column: 13
High quality sequence stop: 692.
FEATURES
1..1039 Location/Qualifiers
source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3872172"
/clone_lib="NIH_MGC_67"

```

```

/tissue_type="retinoblastoma"
/lab_host="PHIOB (phage-resistant)"
/note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dr.
Average insert size 1.75 kb. Library constructed by Life
Technologies."
228 a 351 c 337 g 122 t 1 others
BASE COUNT

```

```

Alignment Scores:
Pred. No.: 1-51e-13 Length: 1039
Score: 296.50 Matches: 95
Percent Similarity: 49.33% Conservatives: 16
Best Local Similarity: 42.22% Mismatches: 66
Query Match: 26.31% Indels: 53
DB: 12 Gaps: 6

US-09-595-947C-10 (1-214) x BE780690 (1-1039)
QY 4 GlnProSer-----GlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
    |||||::: ::::|::|
Db 244 CAGCCGCGCTGAGACGTGCATCTCCGACCTCGACTCGCCACGACGAGCGGCGAGTGAAC 303
QY 21 PheProArgAlaSerGluAspGluValThrCys-----ProThr 33
    ||||| ::::|::| |||
Db 304 TATCCGGCTTCTCTCACG-CAGCAGGAGAGACTGTGCCAGACTCCACAGGACGAGCTCCGCT 362
QY 34 SerAlaProProSerProThrArgThrProGlyAsnCysAlaGluAlaGluGluGly 53
    ||| |||||:::|::| |||
Db 363 TCGGGCGCGCGCGCGCGCGCG----- 386
QY 54 CysArgGlyAlaProArgLysLeuArgAla----- 63
    ||||| |||||
Db 387 ---AGGGCGCGCCCAATATCTCCCGCGCTCTGAGGTTCCAGGGGCGACAGGACGACGAG 443
QY 64 -----ArgArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLys 79
    |||||:::|::|:::|::|
Db 444 CAGAGAGCGCGCGCGCGCGCGCGCGGCGGCTCCGCTCCGAGGCGCTGCGACTCG 503
QY 80 GlnArgSerArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeu 99
    |||||:::|::|:::|::|
Db 504 CTGCGCAGGAGCGCGCGCGCTCAAGGCCAACGATCGCGAGCGCAACGCGCATGCACAACTG 563
QY 100 AsnSerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAlaLysLeu 119
    |||||:::|::|:::|::|
Db 564 AACCGGCGCTTGACGCACTCGCAGCGTCTCGCTCG-TTCCCCGACGAAACCAAGCTC 622
QY 120 ThrLysIleGluThrLeuArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeu 139
    |||||:::|::|:::|::|
Db 623 ACCAAATCAGACGCTGGC-TTCGGCTACAA-TACATCTGGGCTCTGGCGGACACTG 680
QY 140 ArgIleAlaAspHisSerLeuTyrAlaLeuGluProProAlaProHisCysGlyGluLeu 159
    |||||:::|::|:::|::|
Db 681 CGCTGCGGATCAAGGGCTCGCGGAGGGGTGCCGGAAGGCGCTCCTCCGCGGAAG-- 738
QY 160 GlySerProGlyGlyProPro-----GlyAspTrpGlySerLeu 172
    ||||| |||||
Db 739 GGCCTACTCTGGTGGCGGACCCCAAGCGCGCAAAAGAGGGGAGACTGGGCAAGGCA 798
QY 173 TyrSerProValSer-----GlnAlaGlySerLeuSerProAlaAlaSerLeu 188
    ||| ::::|::|
Db 799 CGGGGGCCCCGACATGAGCCACGATCGCGCTCCGAAGAAACAACTAAGCGCGGCGAC 858
QY 189 GluGluArgProGly 193
    |||||
Db 859 CGGACACGACGACGC 873

```

RESULT 7	CNS03VJ9/c	CNS03VJ9	1037 bp	DNA	linear	GSS 18-MAY-2000
LOCUS						
DEFINITION		Tetraodon nigroviridis genome survey sequence T7 end of clone 062M23 of library G from Tetraodon nigroviridis, genomic survey sequence.				

ACCESSION	AL262494
VERSION	AL262494.1 GI:7994120
KEYWORDS	GSS; genome survey sequence.
SOURCE	Tetraodon nigroviridis.
ORGANISM	Tetraodon nigroviridis
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon.
AUTHORS	1 (bases 1 to 1037) Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F., Saurin,W. and Weissenbach,J.
TITLE	Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 1037) Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.
AUTHORS	Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis
TITLE	Unpublished
JOURNAL	3 (bases 1 to 1037) Genoscope.
REFERENCE	Direct Submission
AUTHORS	Submitted (12-APR-2000)
TITLE	This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at
JOURNAL	http://www.genoscope.cns.fr/Tetraodon.
COMMENT	Location/Qualifiers
FEATURES	1..1037 /organism="Tetraodon nigroviridis" /db_xref="taxon:99883" /clone="062423" /clone_lib="G"
source	
BASE COUNT	205 a 254 c 267 g 308 t 3 others
ORIGIN	
Alignment Scores:	
Pred. No.:	1.09e-12 Length: 1037
Score:	285.50 Matches: 66
Percent Similarity:	64.41% Conservative: 10
Best Local Similarity:	55.93% Mismatches: 21
Query Match:	25.33% Indels: 22
DB:	17 Gaps: 4
US-09-595-947C-10 (1-214) x CNS03VJ9 (1-1037)	
Qy 32	ProThrSerAlaProSerProThrArgThrProGlyAsnCysAlaGluAlaGluGlu 51
Db 533	CCCGCGAACCGCTCTCCGGCAGCTCAGATGACACCCC----- 496
Qy 52	GlyGlyCysArgGlyAla-----ProArgLysLeuArgAlaA-ArgGlyGly 67
Db 495	---GGGTCCGGCGACCTTTTGGTAGAGGATCGCGTACAGTAAGA----- 454
Qy 68	ArgSerArgProLysSerGluLeuAlaLeuSerIys-----GlnArgArgSerArgArg 85
Db 453	-----AATCCTTAACATAAAACAGACTTTGAGAAAGTTGGACAGCGGGCCGACGAGG 400
Qy 86	LysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsnSerAlaLeuAspAla 105
Db 399	ATGAAACGGAACGACAGGGAGCGCCACCGGATGCACAACTGAACTCCGCCCTGGAGCGG 340
Qy 106	LeuArgGlyValLeuProThrPheProAspAlaLysLeuThrLysIleGluThrLeu 125
Db 339	CTGAGGGGCGATCTTCGCGGTGCTACCCGAGGACACAAAGCTGACCAAAATTGAACTCTG 280
Qy 126	ArgPheAlaHisAsnTyrlIleTrrPAlaLeuThrGlnThrLeuArgIleAlaAsp 143

ORIGIN

Alignment Scores:		#
Score:	3.48e-10	Length:
DB:	249.00	Matches:
Percent Similarity:	80.82%	Conservative:
Best Local Similarity:	63.01%	Mismatches:
Query Match:	22.09%	Indels:
	10	Gaps:

US-09-595-947C-10 (1-214) x AV995230 (1-490)

68	ArgSerArgProLysSerGluLeuAlaLeuSerLysGlnArgArgSerArgItyLys	87
QY	::: :::	
Db	5 CGCAGACGGGAAGAATCCAGTACAATTAAACAGTTAAAAAATTCGACGATCAAAA	64
QY	88 AlaAsnAspArgGluArgAsnArgMetHisAspLeuAsnSerAlaLeuAspAlaLeuArg	107
Db	65 GCTAATGACAGAGAACCAACCAGATGCATGGCTTAATGACGACACTAGAGAATAAAGA	124
QY	108 GlyValLeuProThrPheProAspAspAlaLysLeuThrLysIleGluThrLeuArgPhe	127
Db	125 CATGTTTTGCCGACCTACCCCGACGAGACCAAATTAACAAAAATCGAAACCTTGCGATTTC	184
QY	128 AlaHisAsnTyrlleTrpalaLeuThrGlnThrLeuArg	140
Db	185 GCATACAAATATATTTGGTCTTTGGAGTAAATCTGTAAAA	223

RESULT 14

BQ178789	814 bp	mRNA	linear	EST 30-APR-2002
UI-M-EVO-bwt-k-07-0-UI.r1	NH-BMAP_EVO	Mus musculus	cdna clone	
IMAGE:5701758	5',	mRNA sequence.		
BQ178789				
BQ178789.1	GI:20354281			
EST.				
house mouse.				
Mus musculus				
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
1 (bases 1 to 814)				
NIH-MGC	http://mgc.nci.nih.gov/ .			
National Institutes of Health, Mammalian Gene Collection (MGC)				
Unpublished (1999)				
Contact: Robert Strausberg, Ph.D.				
Email: cgabps@mail.nih.gov				
Tissue Procurement: Dr. James Lin, University of Iowa				
cdna Library preparation: Dr. M. Bento Soares, University of Iowa				
cdna Library Arrayed by: Dr. M. Bento Soares, University of Iowa				
cdna Sequencing by: Dr. M. Bento Soares, University of Iowa				
Clone Distribution: MGC clone distribution information can be				
found through the I.M.A.G.E. Consortium/LLNL at:				
http://image.llnl.gov				
This clone was contributed by the Brain Molecular Anatomy Project				
(BMAP)				

(BMAF)
Seq primer: NYX-5

```

seq primer: pVX.3.
Location/Qualifiers
1..814
/organism="Mus musculus"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone_image="5701758"
/clone_lib="NIH_BMAP_EV0"
/tissue_type="whole brain"
/dev_stage="embryo 15.5 dpc"
/lab_host="DH10B (T1 phage resistant)"
/note="Organ: brain; Vector: pVX-Asc; Site.1: ECOR I; Site.2: Not I. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured mRNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was size selected according to mRNA size fraction,

```

ligated with EcoR I adaptor, digested with Not I, and then cloned directionally into pX-asc vector. The library tag sequence located between the Not I site and the polyA tail, is GGGCTGGAA. This library was created for the University of Iowa Mouse Brain Molecular Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institutes of Mental Health (NIMH), Hemin Chan, Ph.D., program coordinator." 268 c. 291 g. 102 t. 2 others

BASE COUNT	151 a	268 c	291 a	102 t	2 others
BASE COUNT	151 a	268 c	291 a	102 t	2 others

ORIGIN

Alignment Scores:	
Pred. No.:	1.78e-09
Score:	243.00
Percent Similarity:	52.54%
Best Local Similarity:	41.81%
Query Match:	21.55%
DB:	14
	Gaps: 8
	Indels: 53
	Mismatches: 32
	Conservative: 19
	Matches: 74
	Length: 814

US-09-595-947C-10 (1-214) x B0178789 (1-814)

16	QY	GlutThrGluArgSerPheProArgAla	-----SerGluAspGluVal	29
291	Db	GAGCCGCTAGTAGTGTCTGCAAACTGAGACTCTGGAGTTGAAGAGGAAAGGAGGTA	350	
30	QY	-----ThrCysProThrSerAlaProSerPro	-----	39
351	Db	CTGATGCTGCTGGGCTGGCTTCCCGGCTCGGCGACCTGACCCGATGCTCTCCACG	410	
40	QY	-----ThrArgThrProGlyAsnCys	-----	46
411	Db	GCGGACGAGGAGGACGAGGAGCTGC CGCGCGGGCTCCGCGCGTGGCAGCGTGA	470	
47	QY	AlaGluAlaGluGlu	-----GlyGlyCysArgGly	56
471	Db	GCGGAAGCGGACGAGGGGTGCAGGGCAGTCCGCGCTCGCGGGGTGCGCG	527	
57	QY	AlaProArgGlyLeuArgAla	-----ArgArgGlyGlyArgSerArg	70
528	Db	--CCAGGCGGTGCTGGCGCTGATGCACGAGTGCACGCGTCGCGCGCTCACGG	584	
71	QY	ProLysSerGluLeuAlaLeuSerLysGln	-----ArgArgSerArgArg	85
585	Db	GCGGTCTCCGAGGTGCCANGACGGGGAGAGCGTGTCACGCGATCAAGAAGACCCGCGAG	644	
86	QY	LysLysAlaAsnAspArgGluArgAsnArgMethIsAspLeuAsnSerAlaLeuAspAla	105	
645	Db	CTCAAGGCCAACAAACCGCAGCGCACACCGCATGCACACCTAAACGCCGCTGGACGG	704	
106	QY	LeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThrLysIleGluThrLeu	125	
705	Db	CTGCGGAGGTGCTGCACCTTCGCCGAGGATGCCANGCTCACGAAGATCAGACGCTG	763	
126	QY	ArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArgIleAla	142	
764	Db	CGCTGTCGCGCACATTACATCTGGCGCTCACCGAGACTCTCGCGCTGGCC	814	

RESULTS 15

CNS02IB1 . 1003 bp DNA linear GSS 13-MAY-2000
 Tetraodon nigroviridis genome survey sequence T7 end of clone
 141A20 of library G from Tetraodon nigroviridis, genomic survey
 sequence.
 AL198694
 AL198694.1 GI:7836945
 GSS: genome survey sequence.
 Tetraodon nigroviridis.
 Tetraodon nigroviridis.
 Tetraodon nigroviridis
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
 Tetraodontidae; Tetraodon.
 1 (bases 1 to 1003)
 REFERENCE

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 8, 2003, 11:37:10 ; Search time 4041 seconds
(without alignments)
10514.748 Million cell updates/sec

Title: US-09-595-947C-1
Perfect score: 1460
Sequence: 1 gcaggtagcagagagagcag.....agagtgcctaatccagtgt 1460

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 2054640 seqs, 14551402878 residues

Word size : 8

Total number of hits satisfying chosen parameters: 3419321

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl.*

```

1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.mu.*
20: em.om.*
21: em.or.*
22: em.ov.*
23: em.pat.*
24: em.ph.*
25: em.pl.*
26: em.ro.*
27: em.sts.*
28: em.un.*
29: em.vi.*
30: em.htg_hum.*
31: em.htg_inv.*
32: em.htg_other.*
33: em.htg_mus.*
34: em.htg_pln.*
35: em.htg_rod.*
36: em.htg_mam.*
37: em.htg_vrt.*
38: em.sy.*
39: em.htgo_hum.*
40: em.htgo_mus.*
41: em.htgo_other.*

```

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1460	100.0	1491	6	A91167 Sequence 1
2	1460	100.0	1491	10	Y10619 R.norvegicu
3	634	43.4	166677	2	AC127817 Rattus no
4	92	6.3	861	10	U76208 Mus muscu
5	92	6.3	1861	10	AF364300 Mus muscu
6	92	6.3	5567	10	Y09167 M.musculus
7	92	6.3	138070	2	AC109783 Mus muscu
8	92	6.3	185806	2	AC127417 Mus muscu
9	35	2.4	6123	9	AF303002 Homo sapi
10	35	2.4	91531	2	Continuation (4 of
11	35	2.4	179697	9	AC023886 Homo sapi
12	32	2.2	1330	9	AJ133776 Homo sapi
13	32	2.2	5340	9	AF234829 Homo sapi
14	32	2.2	165110	9	AL450311 Human DNA
15	32	2.2	173341	2	AC021954 Homo sapi
16	26	1.8	170896	2	AC011010 Homo sapi
17	25	1.7	25	6	A91170 Sequence 4
18	23	1.6	735	10	U67776 Mus musculu
19	23	1.6	770	5	AF123884 Gallus ga
20	23	1.6	790	5	AJ012659 Gallus ga
21	23	1.6	932	10	Y09166 M.musculus
22	23	1.6	1315	10	U63841 Mus musculu
23	23	1.6	1333	6	AR023715 Sequence
24	23	1.6	1341	5	AF109014 Gallus ga
25	23	1.6	1385	10	U76207 Mus musculu
26	23	1.6	1412	10	Y07621 M.musculus
27	23	1.6	1880	5	AF303000 Gallus ga
28	23	1.6	10393	10	AF303001 Mus muscu
29	23	1.6	71538	2	AC118243 Mus muscu
30	23	1.6	123855	2	AC102600 Mus muscu
31	23	1.6	174688	2	AC124395 Mus muscu
32	21	1.4	310	6	I84656 Sequence 12
33	21	1.4	1164	5	GGNEURODL
34	21	1.4	1268	6	AR023709 Sequence
35	21	1.4	1268	9	HSU63842 Human neuro
36	21	1.4	1352	6	I84655 Sequence 10
37	21	1.4	1535	6	AR023708 Sequence
38	21	1.4	1550	6	AR103242 Sequence
39	21	1.4	1550	6	BD000105 Different
40	21	1.4	1550	9	HSU58681 Homo sapien
41	21	1.4	1596	5	AF060885 Gallus ga
42	21	1.4	1675	9	BC008687 Homo sapi
43	21	1.4	1717	9	BC028226 Homo sapi
44	21	1.4	1831	10	MUSNDRF
45	21	1.4	1843	10	RAT4

ALIGNMENTS

RESULT 1	A91167	Sequence 1 from Patent	1491 bp	DNA	linear	PAT 22-JAN-2000
LOCUS	A91167	Sequence 1 from Patent	WO9827206.			
DEFINITION	A91167					
ACCESSION	A91167					
VERSION	A91167.1	GI:5740202				
KEYWORDS						
SOURCE						
ORGANISM						
	Rattus sp.					
	Rattus sp.					
	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;					
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;					
	Rattus.					
REFERENCE						
AUTHORS	1 (bases 1 to 1491)					
JOURNAL	Icard-Liepkalns, C., Mallet, J. and Corresponding, N.A.					
	Patent: WO 9827206-A 1 25-JUN-1998;					

ICARD LIEPKALNS CHRISTINE (FR); MALLETT JACQUES (FR)

FEATURES

Location/Qualifiers
 1..1491
 /organism="Rattus sp."
 /db_xref="taxon:10118"
 459..1103
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="CAB69414.1"
 /db_xref="GI:6740203"
 /translation="MAPHPLDAPTIQVSOETQOPFPQASDHEVLSSNPPSPPLVPR
 DCEAEAGDCRGTSRLRARRGGRNRPKSELASKORRRRRKANKRDRNRNHLNSA
 LDALRGVLPFPDDAKLTIETURFAHNYIWAUTQIRADHSFYGPPEPPVPGGELGS
 PGGSSGDWGSYSPVSQAGSLPTASLEFPGLQVPSPLLPTLFLVPSDFL"
 BASE COUNT 307 a 487 c 413 g 284 t
 ORIGIN

Query Match 100.0%; Score 1460; DB 6; Length 1491;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1460; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAGGTAGCAGAGGAGCAGTCCCTGGGCCCCCGCTTGGTGTGATTCGCCCGCTGGCAGACGCA 60
 DB 1 GCAGGTAGCAGAGGAGCAGTCCCTGGGCCCCCGCTTGGTGTGATTCGCCCGCTGGCAGACGCA 60

QY 61 GCAGCCCGGAGGACGCTCCTGTGTCGGGCGAGACAGATAAAGCGTGCCAGGGGACACA 120
 DB 61 GCAGCCCGGAGGACGCTCCTGTGTCGGGCGAGACAGATAAAGCGTGCCAGGGGACACA 120

QY 121 CGATTAGCAGTCAAGAGTCCCTCTGGGTCTCACCCTGCACAGAGCGGAGGACCCCT 180
 DB 121 CGATTAGCAGTCAAGAGTCCCTCTGGGTCTCACCCTGCACAGAGCGGAGGACCCCT 180

QY 181 CCGAGCTCTTTGTGCTCCAGACGCAATTTACTCCAGGCGAGGCGCTGTCAGCTCAG 240
 DB 181 CCGAGCTCTTTGTGCTCCAGACGCAATTTACTCCAGGCGAGGCGCTGTCAGCTCAG 240

QY 241 CAAAGCTTCGAGCAGCAGAGGGGTTCAGTATCCACCGCTGCTTACTCTGACCTGACACCC 300
 DB 241 CAAAGCTTCGAGCAGCAGAGGGGTTCAGTATCCACCGCTGCTTACTCTGACCTGACACCC 300

QY 301 GCAGCTCTCTGTTTGTAGCCCGGAGTAACTAGGTAACTATTTAGGAACCTCCCAAAGGG 360
 DB 301 GCAGCTCTCTGTTTGTAGCCCGGAGTAACTAGGTAACTATTTAGGAACCTCCCAAAGGG 360

QY 361 TAGAAGAGGGAGTGGGTGGCGTACTCTAGTCCCGCTGGAGTGACCTCTAAGTCAGAG 420
 DB 361 TAGAAGAGGGAGTGGGTGGCGTACTCTAGTCCCGCTGGAGTGACCTCTAAGTCAGAG 420

QY 421 ACTGTACACCCCTCCCATTTTCCCAACCTCAGATGGGCGCTCATCCCTTGGATG 480
 DB 421 ACTGTACACCCCTCCCATTTTCCCAACCTCAGATGGGCGCTCATCCCTTGGATG 480

QY 481 CGCCACCATCCAAAGTGTCCCAAGAGACCCAGCAACCTTTCCGGAGCCTCGGACCAAG 540
 DB 481 CGCCACCATCCAAAGTGTCCCAAGAGACCCAGCAACCTTTCCGGAGCCTCGGACCAAG 540

QY 541 AAGTGCTAGTTCOAATTCACCCCACTAGCCCACTCTGTAACCGAGGAGTGTCTCCG 600
 DB 541 AAGTGCTAGTTCOAATTCACCCCACTAGCCCACTCTGTAACCGAGGAGTGTCTCCG 600

QY 601 AAGCAGAGGAGGTGACTCGGAGGGACATCGAGGAGTCCGTCGCGCGGAGGCGG 660
 DB 601 AAGCAGAGGAGGTGACTCGGAGGGACATCGAGGAGTCCGTCGCGCGGAGGCGG 660

QY 661 GCACAGGCGCCCAAGAGGAGTGGCACTGACGACAGCAGCAGAAAGCGGCGCAAGAGG 720
 DB 661 GCACAGGCGCCCAAGAGGAGTGGCACTGACGACAGCAGCAGAAAGCGGCGCAAGAGG 720

QY 721 CCAACGACCGGAGCGCAACCGATGACACACCTTAATCTCCGCGCTGGATGCGCTGCGG 780
 DB 721 CCAACGACCGGAGCGCAACCGATGACACACCTTAATCTCCGCGCTGGATGCGCTGCGG 780

QY 781 GTGTCTGCCCCACCTTCCCGGATGACCCAACTTACAAAGATGAGACCCCTCGCTTCG 840
 DB 781 GTGTCTGCCCCACCTTCCCGGATGACCCAACTTACAAAGATGAGACCCCTCGCTTCG 840

QY 841 CCACAACTACATTTGGGCACTGACTCAGAGCTGCGCATAGCGGACACAGCTTCTACG 900
 DB 841 CCACAACTACATTTGGGCACTGACTCAGAGCTGCGCATAGCGGACACAGCTTCTACG 900

QY 901 GCCCCGAGCCCTGTGCCCTGTGGGAGCTGGGAAGCCGGGAGGGGCTTCCAGCGCGG 960
 DB 901 GCCCCGAGCCCTGTGCCCTGTGGGAGCTGGGAAGCCGGGAGGGGCTTCCAGCGCGG 960

QY 961 ACTGGGGCTCTATCTACTCTCCCAAGTTCCTCCAAAGCTGCTAGCCTGAGCCCAACAGCTCAT 1020
 DB 961 ACTGGGGCTCTATCTACTCTCCCAAGTTCCTCCAAAGCTGCTAGCCTGAGCCCAACAGCTCAT 1020

QY 1021 TGGAGGAGTTCCCTGGGCTGCAGGTGCCAGCTCCCATCTCTGTCTCTCCCGGCAACC 1080
 DB 1021 TGGAGGAGTTCCCTGGGCTGCAGGTGCCAGCTCCCATCTCTGTCTCTCCCGGCAACC 1080

QY 1081 TGGTGTCTCAGACTTCTTGTGAAGGGCCCAACAGAGCCCTGGGCGGTGGGCGCTGGCAG 1140
 DB 1081 TGGTGTCTCAGACTTCTTGTGAAGGGCCCAACAGAGCCCTGGGCGGTGGGCGCTGGCAG 1140

QY 1141 AAAGGAGGAGTGCAGAGTGTCTGAAATGGAAGTGTAGTGAGGAGCTCGAGCATCTCGC 1200
 DB 1141 AAAGGAGGAGTGCAGAGTGTCTGAAATGGAAGTGTAGTGAGGAGCTCGAGCATCTCGC 1200

QY 1201 CCCTTCTGGCTTTCATTAGTCAGTCCCTGATTTAAACAGGATTCGCACAGTTCCTTGCT 1260
 DB 1201 CCCTTCTGGCTTTCATTAGTCAGTCCCTGATTTAAACAGGATTCGCACAGTTCCTTGCT 1260

QY 1261 GCTGTGCTGCACAAAGGACATTCAGGCTGATCTCTTAAACCTCCCTCAGTGTGGCC 1320
 DB 1261 GCTGTGCTGCACAAAGGACATTCAGGCTGATCTCTTAAACCTCCCTCAGTGTGGCC 1320

QY 1321 ACCTCAACTCCCGCTCCAGCAGAGGAGAGCCGTAGCAGCTAAATAGTTGGAGACTCC 1380
 DB 1321 ACCTCAACTCCCGCTCCAGCAGAGGAGAGCCGTAGCAGCTAAATAGTTGGAGACTCC 1380

QY 1381 ATACTTCTGTGACTCGGCCCTCTTCAAAATCTCGGGGCTCCCAACACCGCTTCTCC 1440
 DB 1381 ATACTTCTGTGACTCGGCCCTCTTCAAAATCTCGGGGCTCCCAACACCGCTTCTCC 1440

QY 1441 AGAGTGACCTAAATCCAGTGT 1460
 DB 1441 AGAGTGACCTAAATCCAGTGT 1460

RESULT 2
 RNRELAXT 1491 bp mRNA linear ROD 06-MAY-1997
 LOCUS R.norvegicus mRNA for transcriptional regulator, Relax.
 DEFINITION Y10619
 ACCESSION Y10619
 VERSION Y10619.1 GI:2072737
 KEYWORDS Relax; transcriptional regulator.
 SOURCE Rattus norvegicus.
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 REFERENCE 1 (bases 1 to 1491)
 AUTHORS Ravassard, P., Chatail, F., Mallet, J. and Icard-Liepkalns, C.
 TITLE Relax, a novel rat bHLH transcriptional regulator transiently
 expressed in the ventricular proliferating zone of the developing
 central nervous system
 J. Neurosci. Res. 48 (2), 146-158 (1997)
 JOURNAL 97276390
 MEDLINE 9130143
 PUBMED 2 (bases 1 to 1491)
 REFERENCE Ravassard, P.
 AUTHORS Direct Submission
 TITLE Submitted (20-JAN-1997) P. Ravassard, CNRS UMR 9923, Bat. CERVI,
 JOURNAL

Hopital de la Pitie Salpetriere, 83 Bd. de l'Hopital, F-75013 Paris, FRANCE

Location/Qualifiers
1. .1491

/organism="Rattus norvegicus"
/strain="Wistar"
/db_xref="taxon:10116"
/clone_lib="RT-PCR"
459. 1103
/function="transcriptional regulator"
/codon_start=1
/product="Relax"
/protein_id="CAA71630.1"
/db_xref="GI:2072738"
/db_xref="SPTREMBL:O08718"
/translation="MAPHLDAPTIQVSQTOOPFGASDHEVLSSNTPPSPTLVPR
DCSEAEAGDCTTSKRLKRRGRKPKSELAKRRSRKKKANDRRNRNHLNSA
LDALRGVLPFTDDAKLKIETLRFAHNY IVALTQTLRIADHSFYGPPEVPVPCGELGS
PGGSSGDWGSYISPVSOAGSLSPASLEEFPLQVPSPSCLLPGLTVFSDFL"

BASE COUNT 307 a 487 c 413 g 284 t
ORIGIN

Query Match 100.0%; Score 1460; DB 10; Length 1491;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1460; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAGGTAGCAGAGGAGCAGTCCCTGGGCCCGCCGTTGCTGATTGGCCCGGTGGCAGCAGCA 60
|||||
Db 1 GCAGGTAGCAGAGGAGCAGTCCCTGGGCCCGCCGTTGCTGATTGGCCCGGTGGCAGCAGCA 60
|||||

QY 61 GCAGCCCGCAGCAGCGCTCTGGTCCGGCAGCAGATAGAGCGTCCGAGGGGACACA 120
|||||
Db 61 GCAGCCCGCAGCAGCGCTCTGGTCCGGCAGCAGATAGAGCGTCCGAGGGGACACA 120
|||||

QY 121 CGATTAGCAGCTCAGAAAGTCCCTCTGGGTCTACCACTGCACAGAGCGCGAGACCCCT 180
|||||
Db 121 CGATTAGCAGCTCAGAAAGTCCCTCTGGGTCTACCACTGCACAGAGCGCGAGACCCCT 180
|||||

QY 181 CCGAGTCTCTGTTCTTCTGCTCCAGACGCAATTTACTCCAGCGAGGCGGCTCGAGTCCAG 240
|||||
Db 181 CCGAGTCTCTGTTCTTCTGCTCCAGACGCAATTTACTCCAGCGAGGCGGCTCGAGTCCAG 240
|||||

QY 241 CAAAGCTCGAAGCAGCAGAGGGGTTACGATATCCACCGCTCTTGACTCTGACCAACCC 300
|||||
Db 241 CAAAGCTCGAAGCAGCAGAGGGGTTACGATATCCACCGCTCTTGACTCTGACCAACCC 300
|||||

QY 301 GCAGCTCTCTGTTCTTTTGGAGCCCGGAGTAACCTAGTAACTATTAGGAACCTCCAAAGG 360
|||||
Db 301 GCAGCTCTCTGTTCTTTTGGAGCCCGGAGTAACCTAGTAACTATTAGGAACCTCCAAAGG 360
|||||

QY 361 TAGAAGAGGGAGTGGGTGGGCTACTCTAGTCCCGTGGAGTGACCTCTAAGTCAGAG 420
|||||
Db 361 TAGAAGAGGGAGTGGGTGGGCTACTCTAGTCCCGTGGAGTGACCTCTAAGTCAGAG 420
|||||

QY 421 ACTGTCCACACCCCTTCATTTTCCCAACCTCAGGATGGCGCTCATCCCTTGATG 480
|||||
Db 421 ACTGTCCACACCCCTTCATTTTCCCAACCTCAGGATGGCGCTCATCCCTTGATG 480
|||||

QY 481 CGCCCACTCCAAAGTGTCCCAAGAGACCCAGCAACCCCTTTCCCGGAGCGCTCGAGACACG 540
|||||
Db 481 CGCCCACTCCAAAGTGTCCCAAGAGACCCAGCAACCCCTTTCCCGGAGCGCTCGAGACACG 540
|||||

QY 541 AAGTGTCTAGTTCCTCAATTCACCCACTAGCCCACTCTCTGTAACCGAGGACTGTCTCG 600
|||||
Db 541 AAGTGTCTAGTTCCTCAATTCACCCACTAGCCCACTCTCTGTAACCGAGGACTGTCTCG 600
|||||

QY 601 AGCAGAGCAGGTGACTCCGAGGACATCGAGGAGCTCCGTGCGCGCGGAGGGC 660
|||||
Db 601 AGCAGAGCAGGTGACTCCGAGGAGCATCGAGGAGCTCCGTGCGCGCGGAGGGC 660
|||||

QY 661 GCAACAGGCCAAGAGCGAGTTGGCACTGAGCAAGCAGGACGACGAGCGCGCAAGAG 720
|||||
Db 661 GCAACAGGCCAAGAGCGAGTTGGCACTGAGCAAGCAGGACGACGAGCGCGCAAGAG 720
|||||

QY 721 CCAACGACCGGAGCGCAACCGCATGCGCAACCTTAACCTCCGGCTGGATGCGCTGGCG 780
|||||
Db 721 CCAACGACCGGAGCGCAACCGCATGCGCAACCTTAACCTCCGGCTGGATGCGCTGGCG 780
|||||

QY 781 GTGTCTGCTCCACCTTCCCGGATGAGCCCAAACTTACAAGATCGAGACCTCGCTTCG 840
|||||
Db 781 GTGTCTGCTCCACCTTCCCGGATGAGCCCAAACTTACAAGATCGAGACCTCGCTTCG 840
|||||

QY 841 CCCAACAATACATTTGGGACATGACTCAGACGCTGGCATAGGGACCAAGCTTCTACG 900
|||||
Db 841 CCCAACAATACATTTGGGACATGACTCAGACGCTGGCATAGGGACCAAGCTTCTACG 900
|||||

QY 901 GCCCGAGCGCCCTGTGCCTGTGGGAGCTGGGAAGCCCGGAGGGGCTCCAGCGCG 960
|||||
Db 901 GCCCGAGCGCCCTGTGCCTGTGGGAGCTGGGAAGCCCGGAGGGGCTCCAGCGCG 960
|||||

QY 961 ACTGGGCTCTATCTACTCTCCCACTTCCCAAGCTGTAGCTTCCAGCCACAGCTCAT 1020
|||||
Db 961 ACTGGGCTCTATCTACTCTCCCACTTCCCAAGCTGTAGCTTCCAGCCACAGCTCAT 1020
|||||

QY 1021 TGGAGAGTTCCTCGCCCTGCGAGTCCAGCTCCCACTCTCTCTGCTCCCGGACCC 1080
|||||
Db 1021 TGGAGAGTTCCTCGCCCTGCGAGTCCAGCTCCCACTCTCTCTGCTCCCGGACCC 1080
|||||

QY 1081 TGGTGTCTCAGACTTCTTGTGAAGGCGCCCAACAGCGCCCTGGCGGCTGGCGCTGCAG 1140
|||||
Db 1081 TGGTGTCTCAGACTTCTTGTGAAGGCGCCCAACAGCGCCCTGGCGGCTGGCGCTGCAG 1140
|||||

QY 1141 AAGGAGGAGGATCAGAGCTGTCTGAAGTGAAGGATGAGGAGGCACTCAGCATCTCGC 1200
|||||
Db 1141 AAGGAGGAGGATCAGAGCTGTCTGAAGTGAAGGATGAGGAGGCACTCAGCATCTCGC 1200
|||||

QY 1201 CCCTTCTGCTTTCATTAGTCCCTGATTAAACAGGATTCGCACAGTTCTCTGCT 1260
|||||
Db 1201 CCCTTCTGCTTTCATTAGTCCCTGATTAAACAGGATTCGCACAGTTCTCTGCT 1260
|||||

QY 1261 GCTGTGCTGCACAAAGGACATTCAGGCTGATCTCTTAAACCTCTCAGTGGCC 1320
|||||
Db 1261 GCTGTGCTGCACAAAGGACATTCAGGCTGATCTCTTAAACCTCTCAGTGGCC 1320
|||||

QY 1321 ACCTCAAACTCCCGCTCCAGAGAGGAGCGTAGCAGTAAATAGTTGGGAGACTCCC 1380
|||||
Db 1321 ACCTCAAACTCCCGCTCCAGAGAGGAGCGTAGCAGTAAATAGTTGGGAGACTCCC 1380
|||||

QY 1381 ATACTTCTGCTGACTCCCGCTCTTCAAACTCTGCGGCGCTCCAAACCCAGCTTCTCC 1440
|||||
Db 1381 ATACTTCTGCTGACTCCCGCTCTTCAAACTCTGCGGCGCTCCAAACCCAGCTTCTCC 1440
|||||

QY 1441 AGAGTGACCTAATCCAGTGT 1460
|||||
Db 1441 AGAGTGACCTAATCCAGTGT 1460
|||||

RESULT 3
AC127817 166677 bp DNA linear HTG 19-JUL-2002
Rattus norvegicus clone CH230-259G16, *** SEQUENCING IN PROGRESS
LOCUS
DEFINITION
AC127817
AC127817.1 GI:21908200
VERSION
HTG; HTGS_PHASE1.
KEYWORDS
Rattus norvegicus
SOURCE
Rattus norvegicus
ORGANISM
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
REFERENCE
1 (bases 1 to 166677)
Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-oshan, F.R., Allen, C.,
Alsbrooks, S.L., Amarantunge, H.C., Are, J.R., Ayele, M., Banks, T.,
Barbaria, J., Benton, J., Bimage, K., Blankenburg, K., Bonnin, D.,
Bouck, J., Bowle, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,
Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,

Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,
 Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,
 Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,
 Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,
 Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,
 Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
 Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,
 Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,
 Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
 Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,
 Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,
 Hernandez, O., Hodgson, A., Hoques, M., Holloway, C., Hollins, B.,
 Honsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E.,
 Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,
 Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,
 Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
 Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loulseged, H.,
 Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
 Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,
 Massey, E., Mawhiney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M.,
 Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
 Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
 Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogih, M., Okwuonu, G.,
 Oragunye, N., Oviato, R., Pace, A., Payton, B., Peery, J., Perez, L.,
 Peters, L., Picketts, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,
 Rives, M., Rojas, A., Rojebokan, I., Rolfe, M., Ruiz, S., Savary, G.,
 Scherer, S., Scott, G., Shen, H., Shoohtari, N., Sisson, I.,
 Sodergren, E., Sonalke, T., Sparks, A., Stanley, H., Stone, H.,
 Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
 Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
 Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q.,
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
 Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K.,
 Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
 Weinstock, G. and Gibbs, R.
 Direct Submission
 Unpublished
 2 (bases 1 to 166677)
 Worley, K.C.
 Direct Submission
 Submitted (19-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 ----- Genome Center
 Center: Baylor College of Medicine
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GZXS
 Center clone name: CH230-259G16
 ----- Summary Statistics
 Sequencing vector: Plasmid;
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 73194 bases at least Q40
 Consensus quality: 79767 bases at least Q30
 Consensus quality: 84278 bases at least Q20

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 80 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 1032: contig of 1032 bp in length
 * 1033 1132: gap of unknown length
 * 1133 2146: contig of 1014 bp in length
 * 2147 2246: gap of unknown length

2247
 3350
 3650
 4696
 4796
 6193
 6293
 7968
 8068
 9287
 9387
 10393
 10493
 11603
 11703
 13174
 13274
 14295
 14395
 15983
 16083
 17116
 17216
 18362
 18462
 20372
 20472
 21570
 21670
 22639
 22739
 23089
 24089
 25266
 25366
 26676
 28415
 28515
 29922
 30022
 31952
 32052
 33981
 34081
 35825
 35925
 37573
 39233
 39333
 41121
 41221
 42663
 42763
 44648
 44748
 46212
 46312
 48275
 48376
 49450
 49550
 51299
 51399
 51400
 52750
 52850
 54512
 54612
 55929
 56029
 57532
 57632

3549: contig of 1303 bp in length
 3649: gap of unknown length
 4695: contig of 1046 bp in length
 4795: gap of unknown length
 6192: contig of 1397 bp in length
 6292: gap of unknown length
 7967: contig of 1675 bp in length
 8067: gap of unknown length
 9286: contig of 1219 bp in length
 9386: gap of unknown length
 10392: contig of 1006 bp in length
 10492: gap of unknown length
 11602: contig of 1110 bp in length
 11702: gap of unknown length
 13173: contig of 1471 bp in length
 13273: gap of unknown length
 14294: contig of 1021 bp in length
 14394: gap of unknown length
 15982: contig of 1588 bp in length
 16082: gap of unknown length
 17115: contig of 1033 bp in length
 17215: gap of unknown length
 18361: contig of 1146 bp in length
 18461: gap of unknown length
 20371: contig of 1910 bp in length
 20471: gap of unknown length
 21519: contig of 1048 bp in length
 21619: gap of unknown length
 22638: contig of 1019 bp in length
 22738: gap of unknown length
 23088: contig of 1250 bp in length
 24088: gap of unknown length
 25265: contig of 1177 bp in length
 25365: gap of unknown length
 26675: contig of 1310 bp in length
 28414: contig of 1639 bp in length
 28514: gap of unknown length
 29921: contig of 1407 bp in length
 30021: gap of unknown length
 31951: contig of 1930 bp in length
 32051: gap of unknown length
 33980: contig of 1929 bp in length
 34080: gap of unknown length
 35824: contig of 1744 bp in length
 35924: gap of unknown length
 37572: contig of 1648 bp in length
 37672: gap of unknown length
 39232: contig of 1560 bp in length
 39332: gap of unknown length
 41120: contig of 1788 bp in length
 41220: gap of unknown length
 42662: contig of 1442 bp in length
 42762: gap of unknown length
 44647: contig of 1885 bp in length
 44747: gap of unknown length
 46211: contig of 1464 bp in length
 46311: gap of unknown length
 48275: contig of 1964 bp in length
 48375: gap of unknown length
 49450: contig of 1075 bp in length
 49550: gap of unknown length
 51299: contig of 1749 bp in length
 51399: gap of unknown length
 52749: contig of 1350 bp in length
 52849: gap of unknown length
 54511: contig of 1662 bp in length
 54611: gap of unknown length
 55928: contig of 1317 bp in length
 56028: gap of unknown length
 57531: contig of 1503 bp in length
 57631: gap of unknown length
 59080: contig of 1449 bp in length

```

* 59081 59180: gap of unknown length
* 59181 60513: contig of 1333 bp in length
* 60514 60613: gap of unknown length
* 60614 62621: contig of 2008 bp in length
* 62622 62721: gap of unknown length
* 62722 64707: contig of 1986 bp in length
* 64708 64807: gap of unknown length
* 64808 66176: contig of 1369 bp in length
* 66177 66276: gap of unknown length
* 66277 68684: contig of 2408 bp in length
* 68685 68784: gap of unknown length
* 68785 70841: contig of 2057 bp in length
* 70842 72278: contig of 1337 bp in length
* 72279 73868: gap of unknown length
* 73869 73968: gap of unknown length
* 73969 75979: contig of 2011 bp in length
* 75980 76079: gap of unknown length
* 76080 78068: contig of 1989 bp in length
* 78069 78168: gap of unknown length
* 78169 79454: contig of 1286 bp in length
* 79455 79554: gap of unknown length
* 79555 80709: contig of 1155 bp in length
* 80710 80809: gap of unknown length
* 80810 82754: contig of 1945 bp in length
* 82755 82854: gap of unknown length
* 82855 83396: contig of 2542 bp in length
* 83397 85496: gap of unknown length
* 85497 88065: contig of 2569 bp in length
* 88066 88165: gap of unknown length
* 88166 90986: contig of 2831 bp in length
* 90987 91096: gap of unknown length
* 91097 92577: contig of 1481 bp in length
* 92578 94520: contig of 1843 bp in length
* 94521 94620: gap of unknown length

Query Match
Best Local Similarity 100.0%; Score 634; DB 2; Length 166677;
Matches 634; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 827 GACCTGCTGCTCCGCCACAACTACATTTGGCACTGACTCAGACGCTGCATAGCGGA 886
DB 57641 GACCTGCTGCTCCGCCACAACTACATTTGGCACTGACTCAGACGCTGCATAGCGGA 57700

QY 887 CCACAGCTTACGGCCCGGAGCCCTGTCCTGTCCTGTCGGGAGCTGGGAAGCCCGGGAGG 946
DB 57701 CCACAGCTTACGGCCCGGAGCCCTGTCCTGTCCTGTCGGGAGCTGGGAAGCCCGGGAGG 57760

QY 947 GGGCTCAGGGGCGACTGGGCTCTATCTACTCCCAAGTTTCCCAAGCTGGTAGCCTGAG 1006
DB 57761 GGGCTCAGGGGCGACTGGGCTCTATCTACTCCCAAGTTTCCCAAGCTGGTAGCCTGAG 57820

QY 1007 CCCACAGCCCTCATTGGAGGAGTTCCTGGCTGCTCAGGTGCCAGCTGCCATCCTGCT 1066
DB 57821 CCCACAGCCCTCATTGGAGGAGTTCCTGGCTGCTCAGGTGCCAGCTGCCATCCTGCT 57880

QY 1067 GCTCCCGGGACCCCTGCTGCTCAGACTTCTTGTGAAGGCGCCCAACAGCCCTGGGCG 1126
DB 57881 GCTCCCGGGACCCCTGCTGCTCAGACTTCTTGTGAAGGCGCCCAACAGCCCTGGGCG 57940

QY 1127 GTGGGCGCTGCCAAGAGGAGGAGTACAGCTGTCTGAATGAAGGTAGTAGGCGCA 1186
DB 57941 GTGGGCGCTGCCAAGAGGAGGAGTACAGCTGTCTGAATGAAGGTAGTAGGCGCA 58000

QY 1187 CTCGAGCATCTCGCCCTTCTGGCTTTCATTAGTCAGGTCCCTGATTTAACAGGATTCG 1246
DB 58001 CTCGAGCATCTCGCCCTTCTGGCTTTCATTAGTCAGGTCCCTGATTTAACAGGATTCG 58060

QY 1247 CACAGTTCCTTGCTGCTGTGCGTGCACAAAGGACATTCAGGCTGATCTCTTAACCC 1306
DB 58061 CACAGTTCCTTGCTGCTGTGCGTGCACAAAGGACATTCAGGCTGATCTCTTAACCC 58120

```

```

QY 1307 TCCTCAGTGTGGCCACCTCAAACTCCCGCTCCAGAGCAGAGAGCGGTAGCACTAATA 1366
DB 58121 TCCTCAGTGTGGCCACCTCAAACTCCCGCTCCAGAGCAGAGAGCGGTAGCACTAATA 58180

QY 1367 GTTGGGAGACTCCCATACTTCTCGTGTGACTCCGCGCTCTTTCAAACTCTGCGGGCTCCAA 1426
DB 58181 GTTGGGAGACTCCCATACTTCTCGTGTGACTCCGCGCTCTTTCAAACTCTGCGGGCTCCAA 58240

QY 1427 CCACCGCTTTCTCCAGAGTACCTTAATCCAGTGT 1460
DB 58241 CCACCGCTTTCTCCAGAGTACCTTAATCCAGTGT 58274

RESULT 4
MMU76208
LOCUS Mus musculus neurogenin 3 (ngn3) gene, complete cds.
DEFINITION
ACCESSION U76208
VERSION U76208.1 GI:1815654
KEYWORDS
SOURCE Mus musculus
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 861)
Sommer,L., Ma,Q. and Anderson,D.J.
neurogenins, a novel family of atonal-related bHLH transcription
factors, are putative mammalian neuronal determination genes that
reveal progenitor cell heterogeneity in the developing CNS and PNS
Mol. Cell. Neurosci. 8 (4), 221-241 (1996)
JOURNAL 97153565
MEDLINE 9000438
PUBMED 2 (bases 1 to 861)
Sommer,L., Ma,Q. and Anderson,D.J.
AUTHORS Direct Submission
TITLE Submitted (24-OCT-1996) Biology 216-76, California Institute of
JOURNAL Technology, Howard Hughes Medical Institute, Pasadena, CA 91125,
USA
COMMENT On Feb 5, 1997 this sequence version replaced gi:1666911.
FEATURES
source
1. 861
/organism="Mus musculus"
/db_xref="taxon:10090"
160..804
/gene="ngn3"
160..804
/gene="ngn3"
/function="bHLH transcription factor"
/notes="NGN3"
/codon_start=1
/product="neurogenin 3"
/protein_id="AAC53029.1"
/db_xref="GI:1666912"
DCSEAEEVDCRGTSRLARRGRNRKSELALSQRKRRKKARRRRMRHNLNSA
LDALRGVLPFPDDAKLTKETLRFHNYTWTQTRIDHSFYGPPEVPVPCGELGS
PGSGNGDWGSIYSPVSOAGNLSPASLEFPGLQVPSFYLPLGALVFSDFL"
BASE COUNT 182 a 274 c 250 g 155 t
ORIGIN
Query Match 6.3%; Score 92; DB 10; Length 861;
Best Local Similarity 100.0%; Pred. No. 1.5e-42;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 762 GCGCTGGATGCGCTGCGGGTGTCTGCCACCTTCGCGGATGACGCCAACTTACAAG 821
DB 463 GCGCTGGATGCGCTGCGGGTGTCTGCCACCTTCGCGGATGACGCCAACTTACAAG 522

QY 822 ATCAGAGCCCTGCGCTTCGCCCACTACAT 853
DB 523 ATCAGAGCCCTGCGCTTCGCCCACTACAT 554

```

```

RESULT 5
AF364300
LOCUS Mus musculus neurogenin 3 gene, complete cds. linear DNA 1861 bp ROD 03-MAY-2001
DEFINITION AF364300
ACCESSION AF364300
VERSION AF364300.1 GI:13937128
KEYWORDS Mus musculus.
SOURCE Mus musculus.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 1861)
AUTHORS Lee, J., Smith, S., Watada, H., Lin, J., Scheel, D., Wang, J., Mirmira, R.
and German, M.
TITLE Regulation of the pancreatic pro-endocrine gene neurogenin3
JOURNAL Diabetes (2001) In press
REFERENCE 2 (bases 1 to 1861)
AUTHORS Schwitzgebel, V. and German, M.
TITLE Direct Submission
JOURNAL Submitted (26-MAR-2001) Hormone Research Institute, University of
California San Francisco, 513 Parnassus Ave, HSW1090, San
Francisco, CA 94145-0534, USA
FEATURES
source Location/Qualifiers
1..1861
/organism="Mus musculus"
/db_xref="taxon:10090"
<1093..>1737
/product="neurogenin 3"
1093..1737
/codon_start=1
/product="neurogenin 3"
/db_xref="GI:13937128"
DCSEAEGDCRGTSRKLRARGRGNRPKSELALSKQRSRKKKANDRRNRHNLNSA
LDALRGVLPFPDDAKLTKIETLFAHNYIWAQTOTLRIADHSFYGPPEPPVPCGELGS
PGGSGNDGWSIYSPVQAGNLSPASLEEFPGQVPPSSPYLLPGALVFDFL"

BASE COUNT 397 a 560 c 537 g 367 t
ORIGIN
Query Match 6.3%; Score 92; DB 10; Length 1861;
Best Local Similarity 100.0%; Pred. No. 1.5e-42;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 762 GCGCTGGATGCGCTCGCGGGTGCTCTGCCACCTTCCCGGATGAGCCCAAACTTACAAG 821
DB 1396 GCGCTGGATGCGCTCGCGGGTGCTCTGCCACCTTCCCGGATGAGCCCAAACTTACAAG 1455

QY 822 ATCGAGACCCCTGCGCTTCGCCCAACTACAT 853
DB 1456 ATCGAGACCCCTGCGCTTCGCCCAACTACAT 1487

RESULT 6
MMATH4B
LOCUS M.musculus MATH4B gene. 5567 bp DNA linear ROD 25-OCT-2000
DEFINITION M.musculus MATH4B gene.
ACCESSION Y09167
VERSION Y09167.2 GI:11065737
KEYWORDS bHLH; mammalian atonal homolog 4B; MATH4B; neurogenin 3.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 5567)
AUTHORS Cau, E., Gradwohl, G., Fode, C. and Guilleminot, F.
TITLE Mash1 activates a cascade of bHLH regulators in olfactory neuron
progenitors
JOURNAL Development 124 (8), 1611-1621 (1997)
MEDLINE 97261963
PUBMED 9108377
REFERENCE 2 (bases 1 to 5567)

```

```

AUTHORS Jacquemin, P., Durvieux, S.M., Jensen, J., Godfraind, C., Gradwohl, G.,
Guilleminot, F., Madsen, O.D., Carmeliet, P., Dewerchin, M., Collen, D.,
Rousseau, G.G. and Lemaigre, F.P.
TITLE Transcription factor hepatocyte nuclear factor 6 regulates
pancreatic endocrine cell differentiation and controls expression
of the proendocrine gene ngn3
JOURNAL Mol. Cell. Biol. 20 (12), 4445-4454 (2000)
MEDLINE 20285449
PUBMED 10825208
REFERENCE 3 (bases 1 to 5567)
AUTHORS Gradwohl, G.J.
TITLE Direct Submission
JOURNAL Submitted (04-NOV-1996) G.J. Gradwohl, IGBMC,
CNRS-INSERM-Universite Louis Pasteur, BP163, C.U. de Strasbourg,
F-67404 ILLKIRCH cedex, FRANCE
REMARK Revised by [4]
REFERENCE 4 (bases 1 to 5567)
AUTHORS Lemaigre, F.P.
TITLE Direct Submission
JOURNAL Submitted (01-AUG-2000) Lemaigre F.P., Hormone and Metabolic
Research Unit, Louvain University Medical School, Avenue Hippocrate
75, box 7529, Brussels 1200, BELGIUM
COMMENT On Oct 31, 2000 this sequence version replaced gi:1666087.
FEATURES
Location/Qualifiers
1..5567
/organism="Mus musculus"
/strain="SV129"
/db_xref="taxon:10090"
/clone="lambda GEM12 clone 19C"
/cell_line="D3"
/cell_type="embryonic stem cells"
/clone_lib="Genomic library No.62, MboI partial digests
cloned in BamHI"
/dev_stage="embryos"
958
/bound_moiety="HNF6"
misc_binding 3692
/bound_moiety="HNF6"
misc_binding 4145..5567
gene /gene="MATH4B"
4145..4151
TATA_signal /gene="MATH4B"
/note="putative"
4923..5567
CDS /gene="MATH4B"
/codon_start=1
/product="neurogenin 3"
/protein_id="CAA70366.1"
/db_xref="GI:1668088"
/db_xref="SWISS-PROT:P70661"
/translation="MAPHLDALTIQVSPETQPPGASDHEVLSSNSTPPSPPTLIPR
DCSEAEGDCRGTSRKLRARGRGNRPKSELALSKQRSRKKKANDRRNRHNLNSA
LDALRGVLPFPDDAKLTKIETLFAHNYIWAQTOTLRIADHSFYGPPEPPVPCGELGS
PGGSGNDGWSIYSPVQAGNLSPASLEEFPGQVPPSSPYLLPGALVFDFL"

BASE COUNT 1271 a 1549 c 1564 g 1183 t
ORIGIN
Query Match 6.3%; Score 92; DB 10; Length 5567;
Best Local Similarity 100.0%; Pred. No. 1.4e-42;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 762 GCGCTGGATGCGCTCGCGGGTGCTCTGCCACCTTCCCGGATGAGCCCAAACTTACAAG 821
DB 5226 GCGCTGGATGCGCTCGCGGGTGCTCTGCCACCTTCCCGGATGAGCCCAAACTTACAAG 5285

QY 822 ATCGAGACCCCTGCGCTTCGCCCAACTACAT 853
DB 5286 ATCGAGACCCCTGCGCTTCGCCCAACTACAT 5317

RESULT 7
AC109783/c AC109783 138070 bp DNA linear HTG 07-FEB-2002
LOCUS

```

DEFINITION Mus musculus clone RP23-121F10, WORKING DRAFT SEQUENCE, 17
 unordered pieces,
 AC109783.1
 AC109783.1 GI:18581594
 HTG; HTGS_PHASE1; HTGS_DRAFT.
 KEYWORDS Mus musculus.
 SOURCE Mus musculus.
 ORGANISM Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 138070)
 McCombie,W.R., Baker,J.P., Ballja,V., Dedhia,N.N., de la
 Bastide,M., Katzenberger,F., Kuit,K., King,L., Kirchoff,K.A.,
 Miller,B., Muller,S., Nascimento,L.U., O'Shaughnessy,A.L.,
 Preston,R.R., Santos,L., Spiegel,L.A., Palmer,L., Yang,C. and
 Zukavern,T.
 Mouse Genomic Sequence
 Unpublished
 2 (bases 1 to 138070)
 McCombie,W.R.
 Direct Submission
 Submitted (07-FEB-2002) Lita Annenberg Hazen Genome Sequencing
 Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring
 Harbor, NY 11724, USA
 ----- Genome Center -----
 Center: Lita Annenberg Hazen Genome Center, Cold Spring Harbor
 Laboratory
 Laboratory
 Center code: CSHL
 Web site: <http://www.cshl.org/genseq>
 Contact: mccombie@cshl.org
 ----- Project Information -----
 Project name: RP23-121F10
 Clone name: RP23-121F10
 Insert size: 173000; agarose-fp
 Insert size: 141616; sum-of-contigs
 Quality coverage: 4.00 in Q20 bases; agarose-fp
 Quality coverage: 3.70 in Q20 bases; sum-of-contigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 17 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 17785: contig of 17785 bp in length
 * 17786 17874: gap of unknown length
 * 17875 30356: contig of 12482 bp in length
 * 30357 30444: gap of unknown length
 * 42304 42306: contig of 11862 bp in length
 * 42307 42394: gap of unknown length
 * 42395 53598: contig of 11204 bp in length
 * 53599 53686: gap of unknown length
 * 53687 64356: contig of 10669 bp in length
 * 64356 74016: contig of 9573 bp in length
 * 74017 74104: gap of unknown length
 * 74105 83366: contig of 9262 bp in length
 * 83367 83454: gap of unknown length
 * 83455 92443: contig of 8901 bp in length
 * 92444 100821: contig of 8378 bp in length
 * 100822 100909: gap of unknown length
 * 100910 107529: contig of 6620 bp in length
 * 107530 107617: gap of unknown length
 * 107618 114066: contig of 6449 bp in length
 * 114067 114154: gap of unknown length
 * 114155 118873: contig of 4719 bp in length
 * 118874 118961: gap of unknown length
 * 118962 123619: contig of 4638 bp in length
 * 123620 123707: gap of unknown length
 * 123708 128240: contig of 4533 bp in length
 * 128241 128328: gap of unknown length

* 128329 132682: contig of 4354 bp in length
 * 132683 132770: gap of unknown length
 * 132771 136341: contig of 3571 bp in length
 * 136342 136429: gap of unknown length
 * 136430 138070: contig of 1641 bp in length.
 Location/Qualifiers
 1. 138070
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="RP23-121F10"
 BASE COUNT 35731 a 33657 c 33303 g 33954 t 1425 others
 ORIGIN
 Query Match 6.3% Score 92; DB 2: Length 138070;
 Best Local Similarity 100.0%; Pred. No. 1.2e-42;
 Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 762 GCGCTGGATGCGCTGCGGGTCTCTGCCACCTTCCCGGATGACGCCAACTTCAAAAG 821
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 110917 GCGCTGGATGCGCTGCGGGTCTCTGCCACCTTCCCGGATGACGCCAACTTCAAAAG 110858
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 822 ATCGAGACCTCGGCTTCCGCCACACTACAT 853
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 110857 ATCGAGACCTCGGCTTCCGCCACACTACAT 110826
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 RESULT 8
 AC127417 185806 bp. DNA linear HTG 15-JUL-2002
 LOCUS Mus musculus chromosome UNK clone RP23-459M2, WORKING DRAFT
 DEFINITION SEQUENCE, 36 unordered pieces.
 AC127417
 AC127417.1 GI:21759524
 VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.
 KEYWORDS house mouse.
 SOURCE Mus musculus
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 McPherson,J.D. and Waterston,R.H.
 1 (bases 1 to 185806)
 The sequence of Mus musculus clone
 Unpublished
 JOURNAL
 2 (bases 1 to 185806)
 McPherson,J.D. and Waterston,R.H.
 DIRECT SUBMISSION
 TITLE Direct Submission
 JOURNAL Submitted (15-JUL-2002) Genome Sequencing Center, 4444 Forest Park
 Parkway, St. Louis, MO 63108, USA
 ----- Genome Center -----
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: <http://genome.wustl.edu/gsc/index.shtml>
 Contact: submissions@watson.wustl.edu
 ----- Project Information -----
 Center project name: M_BA0459M02
 ----- Summary Statistics -----
 Sequencing vector: M13; 0%
 Sequencing vector: plasmid; 100%
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 173518 bases at least Q40
 Consensus quality: 177276 bases at least Q30
 Consensus quality: 179611 bases at least Q20
 Insert size: 192000; agarose-fp
 Insert size: 184535; sum-of-contigs
 Quality coverage: 3.66 in Q20 bases; agarose-fp
 Quality coverage: 3.85 in Q20 bases; sum-of-contigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 36 contigs. The true order of the pieces
 * is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

```

1 1042: contig of 1042 bp in length
* 1043
* 1142: gap of unknown length
* 1143
* 2216: contig of 1074 bp in length
* 2217
* 3838: contig of 1522 bp in length
* 3839
* 3939: gap of unknown length
* 3939
* 5396: contig of 1458 bp in length
* 5397
* 5496: gap of unknown length
* 5497
* 6874: contig of 1378 bp in length
* 6875
* 6875: gap of unknown length
* 6875
* 8481: contig of 1507 bp in length
* 8482
* 8581: gap of unknown length
* 8582
* 9874: contig of 1293 bp in length
* 9875
* 9875: gap of unknown length
* 9875
* 11483: contig of 1509 bp in length
* 11484
* 11583: gap of unknown length
* 11584
* 13834: contig of 2251 bp in length
* 13835
* 13934: gap of unknown length
* 13935
* 16035: contig of 2101 bp in length
* 16036
* 16135: gap of unknown length
* 16136
* 19100: contig of 2965 bp in length
* 19101
* 19200: gap of unknown length
* 19201
* 22184: contig of 2984 bp in length
* 22185
* 22284: gap of unknown length
* 22285
* 25287: contig of 3003 bp in length
* 25288
* 25387: gap of unknown length
* 25388
* 28431: contig of 3044 bp in length
* 28432
* 28531: gap of unknown length
* 28532
* 31409: contig of 2878 bp in length
* 31410
* 31509: gap of unknown length
* 31510
* 35320: contig of 3811 bp in length
* 35321
* 35420: gap of unknown length
* 35421
* 38808: contig of 3388 bp in length
* 38809
* 38908: gap of unknown length
* 38909
* 43846: gap of unknown length
* 43847
* 43946: gap of unknown length
* 43947
* 49171: gap of unknown length
* 49172
* 53157: contig of 3886 bp in length
* 53158
* 53257: gap of unknown length
* 53258
* 57731: contig of 4474 bp in length
* 57732
* 57831: gap of unknown length
* 57832
* 62660: contig of 4829 bp in length
* 62661
* 62760: gap of unknown length
* 62761
* 68243: contig of 5483 bp in length
* 68244
* 68343: gap of unknown length
* 68344
* 73047: contig of 4704 bp in length
* 73048
* 73147: gap of unknown length
* 73148
* 78565: contig of 5418 bp in length
* 78566
* 78665: gap of unknown length
* 78666
* 85726: contig of 7061 bp in length
* 85727
* 85826: gap of unknown length
* 85827
* 91985: contig of 6159 bp in length
* 91986
* 92085: gap of unknown length
* 92086
* 97984: contig of 5899 bp in length
* 97985
* 98084: gap of unknown length
* 98085
* 105338: contig of 7254 bp in length
* 105339
* 105439: gap of unknown length
* 105440
* 113368: contig of 7930 bp in length
* 113369
* 113469: gap of unknown length
* 113470
* 120753: contig of 7285 bp in length
* 120754
* 120854: gap of unknown length
* 120855
* 129342: contig of 8489 bp in length
* 129343
* 129442: gap of unknown length
* 129443
* 138998: contig of 9556 bp in length
* 138999
* 139098: gap of unknown length
* 139099
* 149579: contig of 10481 bp in length
* 149580
* 149679: gap of unknown length

```

```

* 149580 162343: contig of 12664 bp in length
* 162344 162443: gap of unknown length
* 162444 185806: contig of 23363 bp in length.
FEATURES
  Location/Qualifiers
    1..185806
      /organism="Mus musculus"
      /db_xref="taxon:10090"
      /chromosome="UNK"
      /clone="RP23-459M2"
    1..1042
      /note="assembly_name:Contig24"
    1143..2216
      /note="assembly_name:Contig25"
    2317..3838
      /note="assembly_name:Contig26"
    3939..5396
      /note="assembly_name:Contig27"
    5497..6874
      /note="assembly_name:Contig28"
    6975..8481
      /note="assembly_name:Contig29"
    8582..9874
      /note="assembly_name:Contig30"
    9975..11483
      /note="assembly_name:Contig31"
    11584..13834
      /note="assembly_name:Contig32"
    13935..16035
      /note="assembly_name:Contig33"
    16136..19100
      /note="assembly_name:Contig34"
    19201..22184
      /note="assembly_name:Contig35"
    22285..25287
      /note="assembly_name:Contig36"
    25388..28431
      /note="assembly_name:Contig37"
    28532..31409
      /note="assembly_name:Contig38"
    31510..35320
      /note="assembly_name:Contig39"
    35421..38808
      /note="assembly_name:Contig40"
    38909..43846
      /note="assembly_name:Contig41"
    43947..49171
      /note="assembly_name:Contig42"
    49272..53157
      /note="assembly_name:Contig43"
    53258..57731
      /note="assembly_name:Contig44"
    57832..62660
      /note="assembly_name:Contig45"
    62761..68243
      /note="assembly_name:Contig46"
    68344..73047
      /note="assembly_name:Contig47"
    73148..78565
      /note="assembly_name:Contig48"
    78666..85726
      /note="assembly_name:Contig49"
    85827..91985
      /note="assembly_name:Contig50"
    92086..97984
      /note="assembly_name:Contig51"
    clone_end:SP6
    vector_side:left
    98085..105338
      /note="assembly_name:Contig52"
    105439..113368
      /note="assembly_name:Contig53"
    113469..120753
      /note="assembly_name:Contig54"

```

misc_feature 120854..129342
/note=assembly_name:Contig55"
misc_feature 129443..138998
/note=assembly_name:Contig56"
misc_feature 139099..149579
/note=assembly_name:Contig57"
misc_feature 149580..162343
/note=assembly_name:Contig58"
misc_feature 162444..185806

Query Match 6.3%; Score 92; DB 2; Length 185806;
Best Local Similarity 100.0%; Pred. No. 1.2e-42;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 762 GCGTGGATGGCTCGCGGTGTCTCGCCACCTTCGCGGATGAGCGCCAACTTACAAG 821
|||||
Db 159127 GCGTGGATGGCTCGCGGTGTCTCGCCACCTTCGCGGATGAGCGCCAACTTACAAG 159186

QY 822 ATCGAGACCTCGCTTCGCCACCAACTACAT 853
|||||
Db 159187 ATCGAGACCTCGCTTCGCCACCAACTACAT 159218

RESULT 9
AF303002 AF303002 6123 bp DNA linear PRI 13-NOV-2001
LOCUS Homo sapiens neurogenin 2 gene, partial cds.
DEFINITION AF303002
ACCESSION AF303002
VERSION AF303002.1 GI:11875763
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 6123)
AUTHORS Simmons,A.D., Horton,S., Abney,A.L. and Johnson,J.E.
TITLE Neurogenin2 expression in ventral and dorsal spinal neural tube
progenitor cells is regulated by distinct enhancers
JOURNAL Dev. Biol. 229 (2), 327-339 (2001)
MEDLINE 21077024
PUBMED 11203697
REFERENCE 2 (bases 1 to 6123)
AUTHORS Simmons,A.D., Horton,S., Abney,A.L. and Johnson,J.E.
TITLE Direct Submission
JOURNAL Submitted (06-SEP-2000) Center for Basic Neuroscience - N44.146, UT
Southwestern Medical Center, 5323 Harry Hines Boulevard, Dallas, TX
75390-9111, USA

FEATURES
Source Location/Qualifiers
1..6123
/organism="Homo sapiens"
/db_xref="taxon:9606"
<1..>754
/product="neurogenin 2"
<1..754
/note="NGN2"
/codon_start=2
/product="neurogenin 2"
/protein_id="AAG40770.1"
/db_xref="GI:11875764"
/translation="ASPAALATPLSSADEEEEPGASGARRQGAEGQGARGG
VAAGEGCPARLLGLVHDCKRRPSRAVSRGAKTAETVQRIKTKRLKANNRRNR
MHNLAALDALREVLPTPEDAKLTETLRFANV IVALTETLADHCGGGGLP
GALFSEAVLLSPGASALSSGDSFSPASTWCTNSPAPSSVSNSTSPISCTLSP
ASPGADMDYWPDPDPKHRYAPHPARDCI"
BASE COUNT 1484 a 1536 c 1507 g 1596 t
ORIGIN

Query Match 2.4%; Score 35; DB 9; Length 6123;
Best Local Similarity 100.0%; Pred. No. 1.7e-08;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 819 AAGATCGAGACCTCGCTTCGCCACCAACTACAT 853
|||||

Db 383 AAGATCGAGACCTCGCTTCGCCACCAACTACAT 417

RESULT 10
AC079846_3/c
WPCOMMENT

Sequence split into 4 fragments LOCUS AC079846 Accession AC079846
Fragment Name Begin End
AC079846_0 1 110000
AC079846_1 100001 210000
AC079846_2 200001 310000
AC079846_3 300001 391531
Continuation (4 of 4) of AC079846 from base 300001 (AC079846 Homo sapiens chromosome)

Query Match 2.4%; Score 35; DB 2; Length 91531;
Best Local Similarity 100.0%; Pred. No. 1.4e-08;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 819 AAGATCGAGACCTCGCTTCGCCACCAACTACAT 853
|||||
Db 86995 AAGATCGAGACCTCGCTTCGCCACCAACTACAT 86961

RESULT 11
AC023886/c
LOCUS AC023886 179697 bp DNA linear PRI 20-MAR-2002
DEFINITION Homo sapiens BAC clone RP11-402J6 from 4, complete sequence.
ACCESSION AC023886
VERSION AC023886.7 GI:19482381
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 179697)
AUTHORS Sulston,J.E. and Waterston,R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
PUBMED 9847074

REFERENCE 2 (bases 1 to 179697)
AUTHORS Radionenko,M. and Abbott,A.
TITLE The sequence of Homo sapiens BAC clone RP11-402J6
JOURNAL Unpublished (2001)
REFERENCE 3 (bases 1 to 179697)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (18-FEB-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

REFERENCE 4 (bases 1 to 179697)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (15-MAR-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

REFERENCE 5 (bases 1 to 179697)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (20-MAR-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Mar 15, 2002 this sequence version replaced gi:17352441.

COMMENT
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@wustl.edu
----- Summary Statistics

Center project name: H_NH0402J06

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping

clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tatenno, M., Catanese, J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org> VECTOR: pBACE3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is AC004049, 2000 bp overlap; the clone sequenced to the right is RP11-148B6. Actual end of this clone is at base position 179697 of RP11-402J6.

FEATURES

```

source
1..179697
/db_xref="taxon:9606"
/map="4"
/clone="RP11-402J6"
/clone_lib="RPCI-11"
1..749
/rpt_family="L1"
734..3067
/rpt_family="L1"
3104..3444
/rpt_family="MaLR"
3447..4289
/rpt_family="L1"
4290..4369
/rpt_family="(TA)n"
4370..4797
/rpt_family="L1"
5022..5194
/rpt_family="L2"
5214..5628
/rpt_family="L1"
5710..5729
/rpt_family="(TA)n"
5749..5863
/rpt_family="(TA)n"
5883..6263
/rpt_family="ERV1"
9901..10151
/rpt_family="ERV1"
10152..10310
/rpt_family="L1"
10311..10400
/rpt_family="Mariner"
10401..10816
/rpt_family="L1"
10967..12362
/rpt_family="L1"
12366..12653
/rpt_family="Alu"

repeat_region 12695..13940
/rpt_family="L1"
repeat_region 13965..14272
/rpt_family="Alu"
repeat_region 14274..14795
/rpt_family="ERV1"
repeat_region 14806..15392
/rpt_family="L1"
repeat_region 21504..22112
/rpt_family="L1"
repeat_region 22131..23166
/rpt_family="L1"
repeat_region 23246..23386
/rpt_family="MaLR"
repeat_region 23511..23654
/rpt_family="MIR"
repeat_region 23655..25486
/rpt_family="L1"
repeat_region 25487..25636
/rpt_family="MIR"
repeat_region 25814..26124
/rpt_family="Alu"
repeat_region 26444..26592
/rpt_family="MIR"
repeat_region 27423..27713
/rpt_family="Alu"
repeat_region 27716..27739
/rpt_family="AT-rich"
repeat_region 27773..27795
/rpt_family="AT-rich"
repeat_region 28020..28429
/rpt_family="ERV1"
repeat_region 29646..29668
/rpt_family="AT-rich"
repeat_region 29727..29871
/rpt_family="L1"
repeat_region 29875..30114
/rpt_family="Alu"
repeat_region 30134..30267
/rpt_family="L1"
repeat_region 30268..30549
/rpt_family="L1"
repeat_region 30550..30883
/rpt_family="MaLR"
repeat_region 32461..32791
/rpt_family="MaLR"
repeat_region 32792..33487
/rpt_family="L1"
repeat_region 33488..33847
/rpt_family="MaLR"
repeat_region 33848..34354
/rpt_family="L1"
repeat_region 34358..34379
/rpt_family="(CATATA)n"
repeat_region 34381..34502
/rpt_family="Alu"
repeat_region 34503..34556
/rpt_family="L1"
repeat_region 34557..34859
/rpt_family="Alu"
repeat_region 34860..35041
/rpt_family="L1"
repeat_region 35219..35676
/rpt_family="L2"
repeat_region 35974..36042
/rpt_family="L2"
repeat_region 36833..37144
/rpt_family="Alu"
repeat_region 36833..36848
/rpt_family="Alu"
misc_feature 37147..37410
/note="match to EST AI537116 (NTID:g4451251) to15h09.x1"
repeat_region 38154..38424
/rpt_family="Alu"

```


Direct Submission
Submitted (12-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
requests: clonerequest@sanger.ac.uk
On Jul 8, 2001 this sequence version replaced gi:14575291.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em:, EMBL; Sw:,
SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C.elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 10, constructed by the Sanger Centre Chromosome 10
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr10
RP11-343J3 is from the library RPCT-11.2 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6
This sequence is the entire insert of clone RP11-343J3 The true
left end of clone RP11-242G30 is at 139955 in this sequence. The
true right end of clone RP11-404C6 is at 6588 in this sequence.

FEATURES
source
1. 165110
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-343J3"
/clone_lib="RPCT-11.2"
7. 147
/note="THE1C repeat: matches 2. 142 of consensus"
repeat_region
151. 293
/note="MIR repeat: matches 2. 148 of consensus"
repeat_region
1136. 1187
/note="26 copies 2 mer gt 98% conserved"
repeat_region
2360. 3272
/note="AluSg repeat: matches 1. 310 of consensus"
repeat_region
4144. 4270
/note="AluSx repeat: matches 1. 134 of consensus"
repeat_region
4319. 4494
/note="AluSx repeat: matches 118. 293 of consensus"
repeat_region
5662. 6438
/note="L1MEC repeat: matches 272. 1095 of consensus"
repeat_region
6916. 8080
/note="L1MEC repeat: matches 1168. 2367 of consensus"
repeat_region
9934. 10087
/note="MIR repeat: matches 25. 184 of consensus"
repeat_region
10353. 10481
/note="MIR repeat: matches 65. 194 of consensus"
repeat_region
11077. 11372
/note="AluSx repeat: matches 1. 294 of consensus"
repeat_region
11600. 11639
/note="20 copies 2 mer tc 95% conserved"
repeat_region
12479. 12608
/note="65 copies 2 mer at 90% conserved"
repeat_region
13433. 13876
/note="MLTLC repeat: matches 1. 466 of consensus"
repeat_region
14372. 14576
/note="AluSg repeat: matches 1. 301 of consensus"
repeat_region
14577. 14633
/note="AluSg repeat: matches 243. 299 of consensus"

```

repeat_region 39214..39423
/Note="MIR repeat: matches 3. .219 of consensus"
repeat_region 39489..439801
/Note="AluSx repeat: matches 1. .292 of consensus"
repeat_region 40050..40189
/Note="MIR repeat: matches 1. .144 of consensus"
repeat_region 41057..41290
/Note="MIR repeat: matches 8. .240 of consensus"
repeat_region 41296..41460
/Note="L2 repeat: matches 2569. .2730 of consensus"
repeat_region 41506..41944
/Note="L2 repeat: matches 1916. .2416 of consensus"
repeat_region 42388..42698
/Note="L2 repeat: matches 1448. .1779 of consensus"
repeat_region 44193..44579
/Note="THEIC repeat: matches 1. .371 of consensus"
repeat_region 44600..44956
/Note="LTR16A repeat: matches 90. .445 of consensus"
repeat_region 45240..45300
/Note="MR58A repeat: matches 37. .97 of consensus"
repeat_region 45798..45909
/Note="LTR41 repeat: matches 90. .192 of consensus"
repeat_region 46826..46871
/Note="23 copies 2 mer gt 97% conserved"
repeat_region 50684..50969
/Note="AluSg repeat: matches 1. .287 of consensus"
repeat_region 50980..51291
/Note="AluSx repeat: matches 1. .312 of consensus"
repeat_region 52222..52519
/Note="AluSx repeat: matches 1. .300 of consensus"
repeat_region 54065..54260
/Note="L1M4 repeat: matches 3865. .4055 of consensus"
repeat_region 54261..54432
/Note="FAM repeat: matches 2. .167 of consensus"
repeat_region 54433..54629
/Note="L1M4 repeat: matches 3652. .3865 of consensus"
repeat_region 54648..54862
/Note="LTR41 repeat: matches 11. .217 of consensus"
repeat_region 54863..55236
/Note="MT1A1 repeat: matches 1. .365 of consensus"
repeat_region 55237..55700
/Note="LTR41 repeat: matches 217. .716 of consensus"

Query Match 2.2%; Score 32; DB 9; Length 165110;
Best Local Similarity 100.0%; Pred. No. 8.6e-07;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 672 AAGAGCGAGTTGGCACTGAGCAAGCAGCGACG 703
|||||
Db 30112 AAGAGCGAGTTGGCACTGAGCAAGCAGCGACG 30081

RESULT 15
AC021954/c 173341 bp DNA linear HTG 24-AUG-2002
LOCUS Homo sapiens chromosome 10 clone RP11-57E12 map 10, WORKING DRAFT
DEFINITION SEQUENCE, 24 unordered pieces.
AC021954
AC021954
AC021954.3 GI:7417809
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 173341)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
Choepey,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
Dearellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
Ferrelia,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Levine,R., Liu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGovern,A., McKernan,K., MCPheeters,R.,
Meldrum,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Olliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Testaye,S., Theodore,J., Tirrell,A., Travers,M., Triggilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,N.

Direct Submission
Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 5, 2000 this sequence version replaced gi:6984451.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L5931
Center clone name: 57.E.12
----- Summary Statistics
Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 161190 bases at least Q40
Consensus quality: 166837 bases at least Q30
Consensus quality: 168995 bases at least Q20
Insert size: 176000; agarose-fp
Insert size: 171041; sum-of-contigs
Quality coverage: 3.7 in Q20 bases; agarose-fp
Quality coverage: 3.8 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 24 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
Direct Submission
Submitted (22-JAN-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 173341)
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepey,Y., Colangelo,M., Collins,S.,
Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferrelia,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehoczyk,J.,
Levine,R., Liu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGovern,A., McKernan,K., MCPheeters,R.,
Meldrum,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Olliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Testaye,S., Theodore,J., Tirrell,A., Travers,M., Triggilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,N.

Direct Submission
Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 5, 2000 this sequence version replaced gi:6984451.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L5931
Center clone name: 57.E.12
----- Summary Statistics
Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 161190 bases at least Q40
Consensus quality: 166837 bases at least Q30
Consensus quality: 168995 bases at least Q20
Insert size: 176000; agarose-fp
Insert size: 171041; sum-of-contigs
Quality coverage: 3.7 in Q20 bases; agarose-fp
Quality coverage: 3.8 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 24 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

```

* 1 1574: contig of 1574 bp in length
* 1575 1674: gap of 100 bp
* 1675 3043: contig of 1369 bp in length
* 3044 3143: gap of 100 bp
* 3144 5889: contig of 2746 bp in length
* 5890 5989: gap of 100 bp
* 5990 8979: contig of 2990 bp in length
* 8980 9079: gap of 100 bp
* 9080 13674: contig of 4595 bp in length
* 13675 13774: gap of 100 bp
* 13775 18831: contig of 5057 bp in length
* 18832 18931: gap of 100 bp
* 18932 23526: contig of 4595 bp in length
* 23527 23626: gap of 100 bp
* 23627 27386: contig of 3760 bp in length
* 27387 27486: gap of 100 bp
* 27487 32572: contig of 5086 bp in length
* 32573 32672: gap of 100 bp
* 32673 38632: contig of 5960 bp in length
* 38633 38732: gap of 100 bp
* 38733 43735: contig of 5003 bp in length
* 43736 43835: gap of 100 bp
* 43836 49020: contig of 5185 bp in length
* 49021 49120: gap of 100 bp
* 49121 53660: contig of 4540 bp in length
* 53661 53760: gap of 100 bp
* 53761 59544: contig of 5784 bp in length
* 59545 59644: gap of 100 bp
* 59645 66395: contig of 6751 bp in length
* 66396 66495: gap of 100 bp
* 66496 74696: contig of 8201 bp in length
* 74697 74796: gap of 100 bp
* 74797 81326: contig of 6530 bp in length
* 81327 81426: gap of 100 bp
* 81427 89446: contig of 8020 bp in length
* 89447 89546: gap of 100 bp
* 89547 96459: contig of 6913 bp in length
* 96460 96559: gap of 100 bp
* 96560 106498: contig of 9939 bp in length
* 106499 106598: gap of 100 bp
* 106599 116986: contig of 10388 bp in length
* 116987 117086: gap of 100 bp
* 117087 128890: contig of 1804 bp in length
* 128891 147290: gap of 100 bp
* 128991 147290: contig of 18300 bp in length
* 147291 147391: gap of 100 bp
* 147391 173341: contig of 25951 bp in length.

```

FEATURES

source

```

1..173341
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/map="10"
/clone="RP11-57E12"
/clone.lib="RPC1-11 Human Male BAC"
1..1574
/note="assembly_fragment"
1675..3043
/note="assembly_fragment"
3144..5889
/note="assembly_fragment"
5990..8979
/note="assembly_fragment"
9080..13674
/note="assembly_fragment"
13775..18831
/note="assembly_fragment"
18932..23526
/note="assembly_fragment"
23627..27386
/note="assembly_fragment"
clone_end:SP6
vector_side:left

```

```

misc_feature 27487..32572
/note="assembly_fragment"
misc_feature 32673..38632
/note="assembly_fragment"
misc_feature 38733..43735
/note="assembly_fragment"
misc_feature 43836..49020
/note="assembly_fragment"
misc_feature 49121..53660
/note="assembly_fragment"
misc_feature 53761..59544
/note="assembly_fragment"
misc_feature 59645..66395
/note="assembly_fragment"
misc_feature 66496..74696
/note="assembly_fragment"
misc_feature 74797..81326
/note="assembly_fragment"
misc_feature 81427..89446
/note="assembly_fragment"
misc_feature 89547..96459
/note="assembly_fragment"
misc_feature 96560..106498
/note="assembly_fragment"
misc_feature 106599..116986
/note="assembly_fragment"
misc_feature 117087..128890
/note="assembly_fragment"
clone_end:17
vector_side:right
misc_feature 128991..147290
/note="assembly_fragment"
misc_feature 147391..173341
/note="assembly_fragment"

BASE COUNT 44173 a 40928 c 40432 g 45501 t 2307 others

Query Match 2.2%; Score 32; DB 2; Length 173341;
Best Local Similarity 100.0%; Pred. No. 8.6e-07;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 672 AAGAGCGAGTTGGCAGTGGAGCAGCAGCGACG 703
Db 6552 AAGAGCGAGTTGGCAGTGGAGCAGCAGCGACG 6521

```

Search completed: April 8, 2003, 20:41:55
Job time : 5789 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 8, 2003, 11:26:20 ; Search time 374 Seconds

(without alignments)

8791.223 Million cell updates/sec

Title: US-09-595-947C-1

Perfect score: 1460

Sequence: 1 gcaggtagcagagagcag.....agagtacctaaccagtgt 1460

Scoring table: OLIGO.NUC

Gapop 60.0 , Gapext 60.0

Searched: 2185239 seqs, 112599159 residues

Word size : 8

Total number of hits satisfying chosen parameters: 2526306

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_101002.*

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description
1	1460	100.0	1491 19 AAV42512	CDNA encoding a no
2	92	6.3	804 19 AAV27050	Mouse neurogenin 3
3	92	6.3	804 21 AAZ51981	Murine neurogenin-1
4	92	6.3	861 22 AAF27266	Murine neurogenin 3
5	92	6.3	1861 21 AAF61090	Murine neurogenin
6	92	6.3	5567 22 AAF27254	Mouse atonal homol
7	65	4.5	65 24 AEN31392	Rat spliced trans
8	35	2.4	65 24 AEN57521	Mouse spliced tran
9	32	2.2	5340 21 AAC61089	Human neurogenin 3

c	10	29	2.0	428	22	AA333797	Human cDNA encodin
	11	26	1.8	26	24	ABT03700	Human Neurogenin-3
	12	25	1.7	25	19	AAV42515	PCR primer used to
	13	23	1.6	738	19	AAV27046	Mouse neurogenin 1
	14	23	1.6	738	21	AAZ51977	Murine neurogenin-
	15	23	1.6	790	22	AAZ51977	Chicken atonal hom
	16	23	1.6	1332	19	AAV42938	DNA encoding murin
	17	23	1.6	1333	18	AAT74894	Mouse neurogenic d
	18	23	1.6	1385	19	AAV27049	Mouse neurogenin 2
	19	23	1.6	1385	21	AAZ51980	Murine neurogenin-
	20	23	1.6	1385	22	AAZ51980	Murine neurogenin 2
	21	23	1.6	1412	22	AAF27269	Mouse atonal homol
	22	23	1.6	1412	22	AAF27255	Mouse atonal homol
	23	21	1.4	1268	18	AAT74891	Human neurogenic d
	24	21	1.4	1268	19	AAV42932	DNA encoding human
	25	21	1.4	1535	18	AAT74890	Human neurogenic d
	26	21	1.4	1535	19	AAV42931	DNA encoding human
	27	21	1.4	1550	21	AA62881	Human NeuroD2 gene
	28	21	1.4	2776	22	AAAL04043	Human reproductive
c	29	21	1.4	2776	22	AAAL04045	Human reproductive
	30	21	1.4	2776	22	AAK68475	Human immune/haema
	31	21	1.4	2776	22	AAK68476	Human immune/haema
c	32	20	1.4	592	24	ABQ49522	Oligonucleotide fo
	33	20	1.4	592	24	ABQ49523	Oligonucleotide fo
c	34	19	1.3	50	17	AAT13245	Acetylcholine tran
c	35	19	1.3	50	17	AAT28933	NACHR beta2 subun
c	36	19	1.3	510	24	ABO91538	M. capsulatus gene
	37	19	1.3	615	24	ABO91539	M. capsulatus gene
c	38	19	1.3	1260	24	ABQ91700	M. capsulatus gene
c	39	19	1.3	1263	21	AA43008	Arabidopsis thalia
c	40	19	1.3	1698	21	AAC36256	Arabidopsis thalia
c	41	19	1.3	3424	22	AAAL06199	Human reproductive
c	42	19	1.3	3424	23	ABU98764	Human testicular a
c	43	19	1.3	3425	22	AAAL06200	Human reproductive
c	44	19	1.3	3425	23	ABU98765	Human testicular a
c	45	19	1.3	18733	22	AAK80682	Human immune/haema

ALIGNMENTS

RESULT 1
AAV42512
ID AAV42512 standard; cDNA; 1491 BP.
XX
AC AAV42512;
XX
DT 05-OCT-1998 (first entry)
XX
DE CDNA encoding a novel BHLH protein designated RELAX.
XX
KW Basic helix-loop-helix; BHLH; RELAX; Rat Embryonic Longitudinal Axis;
KW control; gene expression; transcriptional activator; targeting;
KW protein expression; central nervous system; CNS; treatment;
KW nervous system disorder; ss.
XX
OS Rattus sp.
XX
FH Key
FT CDS
FT Location/Qualifiers
FT 459..1103
FT /*tag= a
FT /product= RELAX
XX
PN WO9827206-A2.
XX
PD 25-JUN-1998.
XX
PF 19-DEC-1997; 97WO-FR02368.
XX
PR 19-DEC-1996; 96FR-0015651.
XX
PA (RHON) RHONE-POULENC RORER SA.
XX

PI Mallet J, Ravassard P, Icard-Liepkalns C;

XX

DR WPI; 1998-362775/
DR P-PSDB: AAW62991.

XX basic helix-loop-helix polypeptide and related nucleic acid - with
PT transcriptional activity, for targeting expression of genes to
PT central nervous system and treatment of nervous disease

XX
PS
Claim 6: Page 20: 28pp: French.

XX The present sequence encodes a basic helix-loop-helix (BHLH) type
CC protein, designated RELAX (Rat Embryonic Longitudinal Axis) protein.
CC The protein is used to control and participate in gene expression,
CC by acting as transcriptional activator, strictly dependent on the
CC presence of an intact E box (CANNTG), particularly for targeting
CC expression of proteins to the central nervous system (CNS). The
CC nucleic acid sequence can be used to treat nervous system disorders,
CC and antisense sequences can be used to control mRNA transcription.
CC

xx
SQ Sequence 1491 BP; 307 A; 487 C; 413 G; 284 T; 0 other;

Query Match	100.0%	Score 1460;	DB 19;	Length 1491;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1460;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY	1	GCAGGTAGCGAGAGGACAGTCCCTGGGCCCCCGTGTCTGATTGGCCCGTGCACAGGCA	60
Db	1	GCAGGTAGCGAGAGGAGACAGTCCCTGGGCCCCCGTGTGATTGGCCCGTGCACAGGCA	60
QY	61	GCAGCCCGGACGACAGCTCCTGGTCCGGGCGACAGACAGATAAAGCGTGCCAGGGGACACA	120
Db	61	GCAGCCCGGACGACAGCTCCTGGTCCGGGCGACAGACAGATAAAGCGTGCCAGGGGACACA	120
QY	121	CGATTAGCAGCTCAGAAGTCCCTCTGGGTCTACCACTGCACAGAGGCGGAGGCCCTT	180
Db	121	CGATTAGCAGCTCAGAAGTCCCTCTGGGTCTACCACTGCACAGAGGCGGAGGCCCTT	180
QY	181	CCGAGCTTCTTGTGCTCCAGACGCAATTACTCCAGCGGAGGCGGCTCGACGTCAG	240
Db	181	CCGAGCTTCTTGTGCTCCAGACGCAATTACTCCAGCGGAGGCGGCTCGACGTCAG	240
QY	241	CAAAACTTCGAAGGAGCAGAGGGGTTACGATATCACCGCTGCTTGACTCTGACACACC	300
Db	241	CAAAACTTCGAAGGAGCAGAGGGGTTACGATATCACCGCTGCTTGACTCTGACACACC	300
QY	301	GCAGCTCTCTGTCTTTTGGAGCCGGAGTAACTAGGTAACTTTAGGAACCTCCAAAGGG	360
Db	301	GCAGCTCTCTGTCTTTTGGAGCCGGAGTAACTAGGTAACTTTAGGAACCTCCAAAGGG	360
QY	361	TAGAAGAGGGAGTGGTGGGCGTACTCTAGTCCC CGCTGGAGTGACCTCTTAAGTCAGAG	420
Db	361	TAGAAGAGGGAGTGGTGGGCGTACTCTAGTCCC CGCTGGAGTGACCTCTTAAGTCAGAG	420
QY	421	ACTGTCAACCCGCTTCCATTTTTTCCCAACCTCAGGATGGCGCTCATCCCTTGGATG	480
Db	421	ACTGTCAACCCGCTTCCATTTTTTCCCAACCTCAGGATGGCGCTCATCCCTTGGATG	480
QY	481	CGCCCAACCTCCAAGTGTCCCAAGAGACCCAGCAACCCCTTCCGGAGGCTCGGACACG	540
Db	481	CGCCCAACCTCCAAGTGTCCCAAGAGACCCAGCAACCCCTTCCGGAGGCTCGGACACG	540
QY	541	AAGTGCTCAGTTCCAATTCCACCCACCTAGCCCCACTCTCGTACCGAGGGACTGCTCGG	600
Db	541	AAGTGCTCAGTTCCAATTCCACCCACCTAGCCCCACTCTCGTACCGAGGGACTGCTCGG	600
QY	601	AAGCAGAAGCAGGTGACTGCGGAGGGACATCGAGGAAGTCCGTGCGCGCGCGAGGGCG	660
Db	601	AAGCAGAAGCAGGTGACTGCGGAGGGACATCGAGGAAGTCCGTGCGCGCGCGAGGGCG	660
QY	661	GCAACAGGCCCAAGAGCGAGTTTGGCTACTGAGCAAGCAGCAGCAAGACGCCGGCGCAAGAG	720
Db	661	GCAACAGGCCCAAGAGCGAGTTTGGCTACTGAGCAAGCAGCAGCAAGACGCCGGCGCAAGAG	720

XX WO9813491-A2.
 PN XX
 XX PR
 PD PD
 XX PD
 XX PF
 XX PF
 XX PR
 PR 17-SEP-1997; 97US-0932411.
 PR 27-SEP-1996; 96US-0722570.
 PR 12-NOV-1996; 96US-0030864.
 PR 19-DEC-1996; 96US-0772009.
 XX
 PA (CALY) CALIFORNIA INST OF TECHNOLOGY.
 XX
 XX PI Anderson DJ, Ma Q, Sommer L;
 XX
 XX WPI; 1998-230702/20.
 DR P-PSDB; AAW54947.
 DR
 XX Mouse neurogenins, useful in neurogenesis - and recombinant nucleic
 XX acids and proteins derived from rat and Xenopus
 PT
 XX
 PS Disclosure; Fig 9; 106pp; English.
 XX
 CC The Mouse neurogenin 3 is one of several neurogenin proteins discussed
 CC in the present invention. The neurogenin nucleic acids can be expressed
 CC in a host cell, transformed using an expression vector, to produce
 CC recombinant proteins. The proteins and the antibodies raised against
 CC the proteins are useful in the study of neurogenesis.
 XX
 XX Sequence 804 BP; 171 A; 263 C; 225 G; 145 T; 0 other;
 SQ

Query Match 6.3%; Score 92; DB 19; Length 804;
 Best Local Similarity 100.0%; Pred. No. 9.3e-35;
 Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 762 GCGCTGGATGCGCTGCGGCTGCTCCGCCACCTTCCCGGATGACGCCAACTTACAAAG 821
 DB 463 GCGCTGGATGCGCTGCGGCTGCTCCGCCACCTTCCCGGATGACGCCAACTTACAAAG 522

QY 822 ATCGAGACCTCGCTTCGCCCAACTACAT 853
 DB 523 ATCGAGACCTCGCTTCGCCCAACTACAT 554

RESULT 3
 AAZ51981
 ID AAZ51981 standard; DNA; 804 BP.
 AC AAZ51981;
 XX
 XX 04-JUL-2000 (first entry)
 DT
 DE Murine neurogenin-3 (NGN3) nucleic acid sequence.
 XX
 XX Neurogenin-3; NGN-3; non-neuronal cell; NNC; neurogenesis;
 KW Phoxa protein; neuronal subtype-specific marker; growth factor;
 KW neural differentiation; transplantation; neuronal dysfunction;
 KW optical nerve damage; auditory nerve damage, neurodegenerative disorder;
 KW neuroprotective; nootropic; anticonvulsant; antiparkinsonian; vulnary;
 KW cerebroprotective; immunosuppressant; antinfecious; ds.
 XX
 OS Mus sp.
 XX
 XX Key Location/Qualifiers
 FH 160..804
 FT CDS
 FT /*tag- a
 FT /product- "Murine neurogenin-3 protein"
 XX
 XX WO200009676-A2.
 PN XX
 XX PD
 PD 24-FEB-2000.
 XX

PF 13-AUG-1999; 99WO-US18525.
 XX
 PR 14-AUG-1998; 98US-0096630.
 XX
 PA (CALY) CALIFORNIA INST OF TECHNOLOGY.
 XX
 PI Anderson DJ, Lo L;
 XX
 XX WPI; 2000-256250/22.
 DR P-PSDB; AAY70570.
 XX
 PT Inducing non-neuronal cells to differentiate into neurons and for
 PT non-neuronal cells to express a neuronal subtype-specific marker,
 PT comprising contacting the non-neuronal cells with a vector containing
 PT neurogenin nucleic acid -
 XX
 PS Claim 1; Fig 1J; 76pp; English.
 XX
 CC The patent discloses a method for inducing non-neuronal cells (NNC) to
 CC differentiate into neurons and for NNCs to express a neuronal subtype
 CC -specific marker. Transformed host cells are used as sources of neuronal
 CC and other growth factors; in culture for screening compounds that
 CC modulate neural differentiation or as sources of recombinantly produced
 CC neurogenins and Phoxa proteins for use in transplantation. The cells
 CC also have a variety of in vivo uses, e.g. for transplantation at sites of
 CC neuronal dysfunction e.g. patients with hearing or vision loss due to
 CC optical or auditory nerve damage, brain or spinal cord injuries, and
 CC neurodegenerative disorders e.g. Alzheimer's disease. The present
 CC sequence encodes murine neurogenin-3 (NGN-3), a transcription factor.
 CC NNCs differentiate into neurons through the recombinant expression of a
 CC transcription factor that induces a core program of neurogenesis. Forced
 CC expression of murine NGN3 can elicit expression of at least some neuronal
 CC phenotypic markers even in NNCs.
 XX
 SQ Sequence 804 BP; 171 A; 263 C; 225 G; 145 T; 0 other;
 SQ

Query Match 6.3%; Score 92; DB 21; Length 804;
 Best Local Similarity 100.0%; Pred. No. 9.3e-35;
 Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 762 GCGCTGGATGCGCTGCGGCTGCTCCGCCACCTTCCCGGATGACGCCAACTTACAAAG 821
 DB 463 GCGCTGGATGCGCTGCGGCTGCTCCGCCACCTTCCCGGATGACGCCAACTTACAAAG 522

QY 822 ATCGAGACCTCGCTTCGCCCAACTACAT 853
 DB 523 ATCGAGACCTCGCTTCGCCCAACTACAT 554

RESULT 4
 AAP27266
 ID AAP27266 standard; cDNA; 861 BP.
 XX
 XX AAP27266;
 XX
 XX 24-APR-2001 (first entry)
 DT
 XX Mouse neurogenin 3 (ngn3) cDNA, SEQ ID NO:24.
 DE
 XX Atonal; homologue; orthologue; atonal-associated protein; deafness;
 KW hearing impairment; vestibular effect; balance disorder; osteoarthritis;
 KW cellular proliferation; cerebellar granule neuron; gene therapy;
 KW mechanoreceptive cell growth; auditory; osteopathic; cytostatic;
 KW transgenic animal; ss.
 XX
 OS Mus musculus.
 XX
 XX WO200073764-A2.
 PN
 XX
 XX 07-DEC-2000.
 PD
 XX
 PF 01-JUN-2000; 2000WO-US15410.
 XX

PR 01-JUN-1999; 99US-0137060.
 PR 19-JAN-2000; 2000US-0176993.
 PA (BAYU) BAYLOR COLLEGE MEDICINE.
 PI Zoghbi HY, Bellen H, Birmingham N, Hassan B, Ben-Arie N;
 XX WPI; 2001-032190/04.
 DR P-PSDB; AAB60359.
 XX
 XX Therapeutic use of atonal-associated nucleic acids or amino acids, or
 PT any of its homologs or orthologs, for the treatment of e.g. deafness,
 PT osteoarthritis and abnormal cell proliferation -
 XX
 PS Disclosure; Page -: 142pp; English.
 XX
 CC The invention relates to the use of atonal-associated nucleic acid or
 CC amino acid sequence, or any of its homologues or orthologues as
 CC therapeutic agents for the treatment of deafness, partial hearing loss,
 CC vestibular effects due to damage or loss of inner hair cells,
 CC osteoarthritis and abnormal cell proliferation. The invention also
 CC encompasses methods of screening for compounds which affect the
 CC expression of an atonal-associated nucleic acid sequence in an animal,
 CC and a transgenic animal in which an allele of a native atonal-associated
 CC gene is replaced by a heterologous nucleic acid sequence, thus
 CC inactivating the atonal-associated allele. The nucleic acids or proteins
 CC may be used in a method of treating an animal for hearing impairment,
 CC joint disease, balance disorders, abnormal cell proliferation, or other
 CC disease related to loss of a functional atonal-associated nucleic acid or
 CC protein. They may particularly be used to treat an animal with a
 CC deficiency in cerebellar granule neurons or their precursors, and may
 CC also be used in promoting mechanoreceptive cell growth and generating
 CC hair cells. The present sequence represents an atonal-associated nucleic
 CC acid sequence referred to in the invention.
 CC Note: The present sequence is not shown in the specification, but
 CC was obtained from GenBank.
 XX
 SQ Sequence 861 BP; 182 A; 274 C; 250 G; 155 T; 0 other;

Query Match 6.3%; Score 92; DB 22; Length 861;
 Best Local Similarity 100.0%; Pred. No. 9.3e-35;
 Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 762 GCGGTGATGCGTGCCTGCCACCTTCCTCCGGATGACGCCAACTTACAAAG 821
 Db 463 GCGGTGATGCGTGCCTGCCACCTTCCTCCGGATGACGCCAACTTACAAAG 522

QY 822 ATCGAGACCTGCGCTTCGCCCAACTACAT 853
 Db 523 ATCGAGACCTGCGCTTCGCCCAACTACAT 554

RESULT 5
 AAC61090
 ID AAC61090 standard; DNA; 1861 BP.
 XX
 AC AAC61090;
 XX
 XX 05-FEB-2001 (first entry)
 XX
 DE Murine neurogenin 3 (Ngn3) genomic DNA sequence.
 XX
 KW Neurogenin 3; Ngn3; cellular differentiation; diabetes mellitus;
 KW islet cell precursor identification; mouse; ds.
 XX
 OS Mus musculus.

Key Location/Qualifiers
 CDS 1093..1737
 FT /*tag= a
 FT /product= "Ngn3"
 FT /note= "Neurogenin 3"
 XX

PN WO200059936-A1.
 XX 12-OCT-2000.
 XX 28-MAR-2000; 2000WO-US08436.
 XX 06-APR-1999; 99US-0128180.
 XX (REGC) UNIV CALIFORNIA.
 XX German MS, Lin J;
 XX WPI; 2000-664989/64.
 DR P-PSDB; AAY85618.
 XX
 XX Novel human neurogenin 3 polypeptides and polynucleotides encoding
 PT them, useful for diagnosis, prevention and treatment of diabetes
 PT mellitus and to identify individuals at risk of diabetes -
 XX
 PS Claim 18; Page 49-50; 54pp; English.
 XX
 CC The human neurogenin 3 Ngn3 DNA sequence AAC61089 encodes the Ngn3
 CC protein AAY85617. The Ngn3 gene is located at chromosome position
 CC 10q22.1-22.2. The invention relates to the human Ngn3 nucleotide and
 CC protein sequences, and includes an antibody recognising the Ngn3 protein.
 CC Also included in the invention is a method for identifying an islet cell
 CC precursor, the method involves analysing a cell for the expression of the
 CC Ngn3 gene product, where detection of the product is indicative of an
 CC islet cell precursor. The Ngn3 DNA sequence is useful as a diagnostic
 CC reagent for detecting (in a subject) a predisposition to a defect in
 CC pancreatic islet cell function or formation associated with a defect in
 CC Ngn3 activity. The Ngn3 protein is useful for identifying beta-cell
 CC precursor cells expressing Ngn3, and to alter cellular differentiation in
 CC culture in vivo to produce new beta-cells to treat patients with diabetes
 CC mellitus. The present sequence represents the murine Ngn3 genomic DNA
 CC sequence.
 XX
 SQ Sequence 1861 BP; 397 A; 560 C; 537 G; 367 T; 0 other;

Query Match 6.3%; Score 92; DB 21; Length 1861;
 Best Local Similarity 100.0%; Pred. No. 9e-35;
 Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 762 GCGGTGATGCGTGCCTGCCACCTTCCTCCGGATGACGCCAACTTACAAAG 821
 Db 1396 GCGGTGATGCGTGCCTGCCACCTTCCTCCGGATGACGCCAACTTACAAAG 1455

QY 822 ATCGAGACCTGCGCTTCGCCCAACTACAT 853
 Db 1456 ATCGAGACCTGCGCTTCGCCCAACTACAT 1487

RESULT 6
 AAF27254
 ID AAF27254 standard; cDNA; 5567 BP.
 XX
 AC AAF27254;
 XX
 XX 24-APR-2001 (first entry)
 XX
 DE Mouse atonal homologue 5 (ATO5, Math4B) cDNA, SEQ ID NO:4.

XX Atonal; homologue; orthologue; atonal-associated protein; deafness;
 KW hearing impairment; vestibular effect; balance disorder; osteoarthritis;
 KW cellular proliferation; cerebellar granule neuron; gene therapy;
 KW mechanoreceptive cell growth; auditory; osteopathic; cytostatic;
 KW transgenic animal; ss.

OS Mus musculus.
 XX
 XX WO200073764-A2.
 XX 07-DEC-2000.

XX 01-JUN-2000; 2000WO-US15410.
 XX PF
 XX 01-JUN-1999; 99US-0137060.
 PR 19-JAN-2000; 2000US-0176993.
 XX PA
 XX (BAYU) BAYLOR COLLEGE MEDICINE.
 XX PI Zoghbi HY, Bellen H, Birmingham N, Hassan B, Ben-Arie N;
 XX DR
 XX WPI; 2001-032190/04.
 DR P-PSDB; AAB60350.
 XX
 XX Therapeutic use of atonal-associated nucleic acids or amino acids, or
 PT any of its homologs or orthologs, for the treatment of e.g. deafness,
 PT osteoarthritis and abnormal cell proliferation -
 XX
 XX Disclosure; Page -; 142pp; English.
 XX
 XX The invention relates to the use of atonal-associated nucleic acid or
 CC amino acid sequence, or any of its homologues or orthologues as
 CC therapeutic agents for the treatment of deafness, partial hearing loss,
 CC vestibular effects due to damage or loss of inner hair cells,
 CC osteoarthritis and abnormal cell proliferation. The invention also
 CC encompasses methods of screening for compounds which affect the
 CC expression of an atonal-associated nucleic acid sequence in an animal,
 CC and a transgenic animal in which an allele of a native atonal-associated
 CC gene is replaced by a heterologous nucleic acid sequence, thus
 CC inactivating the atonal-associated allele. The nucleic acids or proteins
 CC may be used in a method of treating an animal for hearing impairment,
 CC joint disease, balance disorders, abnormal cell proliferation, or other
 CC disease related to loss of a functional atonal-associated nucleic acid or
 CC protein. They may particularly be used to treat an animal with a
 CC deficiency in cerebellar granule neurons or their precursors, and may
 CC also be used in promoting mechanoreceptive cell growth and generating
 CC hair cells. The present sequence represents an atonal-associated nucleic
 CC acid sequence referred to in the invention.
 CC Note: The present sequence is not shown in the specification, but
 CC was obtained from Genbank.
 XX
 XX Sequence 5567 BP; 1271 A; 1549 C; 1564 G; 1183 T; 0 other;
 SQ
 Query Match 6.3%; Score 92; DB 22; Length 5567;
 Best Local Similarity 100.0%; Pred. No. 8.5e-35;
 Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 762 GCCTGGATGCGTGGCGGTGTCCTGCCACCTTCCCGGATGACGCCAACTTACAAG 821
 DB 5226 GCCTGGATGCGTGGCGGTGTCCTGCCACCTTCCCGGATGACGCCAACTTACAAG 5285
 QY 822 ATCGAGACCTGGCTGCGCCCACTACAT 853
 DB 5286 ATCGAGACCTGGCTGCGCCCACTACAT 5317
 RESULT 7
 ABN31392
 ID ABN31392 standard; DNA; 65 BP.
 XX
 XX ABN31392;
 XX
 XX 15-JUL-2002 (first entry)
 XX
 XX Rat spliced transcript detection oligonucleotide SEQ ID NO:4140.
 DE Human; mouse; rat; splice transcript; detection; RNA transcript;
 KW splice variant; transcriptome; oligonucleotide library; ss.
 XX
 XX Rattus norvegicus.
 OS
 XX WO20021049-A2.
 XX
 XX 07-FEB-2002.
 PD

XX 20-JUL-2001; 2001WO-IB01903.
 XX PF
 XX 28-JUL-2000; 2000US-221607P.
 PR 02-MAY-2001; 2001US-287724P.
 XX PA
 XX (COMP-) COMPUGEN INC.
 XX PI Shoshan A, Wasserman A, Mintz E, Mintz L, Faigler S;
 XX DR
 XX WPI; 2002-257383/30.
 XX
 XX New oligonucleotide libraries comprising oligonucleotides which
 PT selectively hybridize to mRNAs transcribed from a transcription unit of
 PT a genome, useful for detecting tissue-, pathology-, and
 PT developmental-specific genes -
 XX
 XX Example 1; SEQ ID 4140; 47pp; English.
 XX
 XX The present invention describes oligonucleotide libraries for detecting
 CC messenger RNAs that populate a (sub-)transcriptome, where the
 CC (sub-)transcriptome comprises messenger RNAs transcribed from multiple
 CC transcription units that populate a genome. The library comprises
 CC several oligonucleotides, each capable of hybridising selectively to a
 CC set of messenger RNAs transcribed from a given transcription unit of
 CC the genome, which encodes one or more messenger RNA splice variants.
 CC The oligonucleotide libraries are useful for detecting mRNAs from a
 CC biological sample, in expression profiling studies, in qualitatively or
 CC quantitatively characterising the corresponding transcriptome, and in
 CC detecting RNA transcripts and splice variants of human or animal
 CC transcriptomes. The libraries may also be used as specialised mini
 CC libraries to detect transcripts of a sub-transcriptome under a
 CC particular biological or pathological state, and so allowing the
 CC detection of tissue- and pathology-specific genes such as those genes
 CC only expressed in specific tissue under a specific pathological
 CC condition; to detect developmental specific genes; and to detect RNA
 CC transcripts and splice variants of a transcriptome of a patient suffering
 CC from a particular disorder. ABN27253 to ABN59589 represent
 CC oligonucleotide sequences from rats, humans and mice, which are used in
 CC the exemplification of the present invention.
 CC N.B. The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 XX Sequence 65 BP; 11 A; 20 C; 14 G; 20 T; 0 other;
 SQ
 Query Match 4.5%; Score 65; DB 24; Length 65;
 Best Local Similarity 100.0%; Pred. No. 1.8e-21;
 Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1252 TTCCTTGCTGCTGGCGTGCACAAAGGACATTCGAGCTGATCCTTAACCCCTC 1311
 DB 1 TTCCTTGCTGCTGGCGTGCACAAAGGACATTCGAGCTGATCCTTAACCCCTC 60
 QY 1312 AGTGT 1316
 DB 61 AGTGT 65
 RESULT 8
 ABN57521
 ID ABN57521 standard; DNA; 65 BP.
 XX
 XX ABN57521;
 XX
 XX 15-JUL-2002 (first entry)
 XX
 XX Mouse spliced transcript detection oligonucleotide SEQ ID NO:30269.
 DE Human; mouse; rat; splice transcript; detection; RNA transcript;
 KW splice variant; transcriptome; oligonucleotide library; ss.
 XX
 XX Mus musculus.
 OS

```

XX PN WO200210449-A2.
XX XX
XX PD 07-FEB-2002.
XX XX
XX PF 20-JUL-2001; 2001WO-IB01903.
XX XX
XX PR 28-JUL-2000; 2000US-221607P.
XX PR 02-MAY-2001; 2001US-287724P.
XX XX
XX FA (COMP-) COMPUGEN INC.
XX XX
XX PI Shoshan A, Wasserman A, Mintz E, Mintz L, Faigler S;
XX XX
XX DR WPI; 2002-257383/30.
XX XX
XX PT New oligonucleotide libraries comprising oligonucleotides which
XX PT selectively hybridize to mRNAs transcribed from a transcription unit of
XX PT a genome, useful for detecting tissue-, pathology-, and
XX PT developmental-specific genes.
XX XX
XX PS Example 1; SEQ ID 30269; 47pp; English.
XX XX
XX CC The present invention describes oligonucleotide libraries for detecting
XX CC messenger RNAs that populate a (sub-)transcriptome, where the
XX CC (sub-)transcriptome comprises messenger RNAs transcribed from multiple
XX CC transcription units that populate a genome. The library comprises
XX CC several oligonucleotides, each capable of hybridising selectively to a
XX CC set of messenger RNAs transcribed from a given transcription unit of
XX CC the genome, which encodes one or more messenger RNA splice variants.
XX CC The oligonucleotide libraries are useful for detecting mRNAs from a
XX CC biological sample, in expression profiling studies, in qualitatively or
XX CC quantitatively characterising the corresponding transcriptome, and in
XX CC detecting RNA transcripts and splice variants of human or animal
XX CC transcriptomes. The libraries may also be used as specialised mini
XX CC libraries to detect transcripts of a sub-transcriptome under a
XX CC particular biological or pathological state, and so allowing the
XX CC detection of tissue- and pathology-specific genes such as those genes
XX CC only expressed in specific tissue under a specific pathological
XX CC condition; to detect developmental specific genes; and to detect RNA
XX CC transcripts and splice variants of a transcriptome of a patient suffering
XX CC from a particular disorder. ABN27253 to ABN59589 represent
XX CC oligonucleotide sequences from rats, humans' and mice, which are used in
XX CC the exemplification of the present invention.
XX CC N.B. The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pot_sequences.
XX XX
XX SQ Sequence 65 BP; 16 A; 25 C; 11 G; 13 T; 0 other;

Query Match 2.4%; Score 35; DB 24; Length 65;
Best Local Similarity 100.0%; Pred. No. 8.8e-07;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 547 TCAGTTCCTCAATCCACCCACCTAGCCCACTCTC 581
Db 1 TCAGTTCCTCAATCCACCCACCTAGCCCACTCTC 35
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

RESULT 9
AAC61089
ID AAC61089 standard; DNA; 5340 BP.
XX XX
XX AC AAC61089;
XX XX
XX DT 05-FEB-2001 (first entry)
XX XX
XX DE Human neurogenin 3 (Ngn3) genomic DNA sequence.
XX XX
XX KW Neurogenin 3; Ngn3; chromosome 10q22.1-22.2; cellular differentiation;
XX KW islet cell precursor identification; diabetes mellitus; human; ds.
XX OS Homo sapiens.

```

```

XX FH Key Location/Qualifiers
XX FT CDS 3022..3666
XX FT /*tag= a
XX FT /*product= "Ngn3"
XX FT /*note= "Neurogenin 3"
XX PN WO200059936-A1.
XX XX
XX PD 12-OCT-2000.
XX XX
XX PF 28-MAR-2000; 2000WO-US08436.
XX XX
XX PR 06-APR-1999; 99US-0128180.
XX XX
XX FA (REGC ) UNIV CALIFORNIA.
XX XX
XX PI German MS, Lin J;
XX XX
XX DR WPI; 2000-664989/64.
XX DR P-PSDB; AAY85617.
XX XX
XX PT Novel human neurogenin 3 polypeptides and polynucleotides encoding
XX PT them, useful for diagnosis, prevention and treatment of diabetes
XX PT mellitus and to identify individuals at risk of diabetes.
XX XX
XX PS Claim 6; Page 46-48; 54pp; English.
XX XX
XX CC The human neurogenin 3 Ngn3 DNA sequence AAC61089 encodes the Ngn3
XX CC protein AAY85617. The Ngn3 gene is located at chromosome position
XX CC 10q22.1-22.2. The invention relates to the human Ngn3 nucleotide and
XX CC protein sequences, and includes an antibody recognising the Ngn3 protein.
XX CC Also included in the invention is a method for identifying an islet cell
XX CC precursor, the method involves analysing a cell for the expression of the
XX CC Ngn3 gene product, where detection of the product is indicative of an
XX CC islet cell precursor. The Ngn3 DNA sequence is useful as a diagnostic
XX CC reagent for detecting (in a subject) a predisposition to a defect in
XX CC pancreatic islet cell function or formation associated with a defect in
XX CC Ngn3 activity. The Ngn3 protein is useful for identifying beta-cell
XX CC precursor cells expressing Ngn3, and to alter cellular differentiation in
XX CC culture in vivo to produce new beta-cells to treat patients with diabetes
XX CC mellitus.
XX XX
XX SQ Sequence 5340 BP; 1215 A; 1500 C; 1514 G; 1111 T; 0 other;

Query Match 2.2%; Score 32; DB 21; Length 5340;
Best Local Similarity 100.0%; Pred. No. 2.1e-05;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 672 AAGACGCGAGTTGGCACTGACGAAGCAGCGACG 703
Db 3235 AAGACGCGAGTTGGCACTGACGAAGCAGCGACG 3266
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

RESULT 10
AAS33797
ID AAS33797 standard; cDNA; 428 BP.
XX XX
XX AC AAS33797;
XX XX
XX DT 17-DEC-2001 (first entry)
XX XX
XX DE Human cDNA encoding a novel foetal antigen, SEQ ID No 321.
XX XX
XX KW Human; foetal tissue antigen; ss; antiinflammatory; neuroprotective;
XX KW immunomodulator; cardiovascular; cytostatic; nephrothropic;
XX KW cardiovascular; autoimmune disease; rheumatoid arthritis;
XX KW hyperproliferative disorder; breast neoplasm; cancer;
XX KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
XX KW cerebral ischaemia; angiogenesis; nervous system disorder;
XX KW Alzheimer's disease; infection; ocular disorder; corneal infection;
XX KW wound healing; epithelial cell proliferation; food additive.
XX XX

```

OS Homo sapiens.
XX WO200155312-A2.
XX PD 02-AUG-2001.
XX PF 17-JAN-2001; 2001WO-US01321.
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0198874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0228279.
PR 22-AUG-2000; 2000US-0228681.
PR 22-AUG-2000; 2000US-0228688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 21-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 11-DEC-2000; 2000US-0251990.
PR 05-JAN-2001; 2001US-0259678.
XX
XX
(HUMA-) HUMAN GENOME SCI INC.
PA

XX Rosen CA, Barash SC, Ruben SM;
 XX WPI; 2001-488782/53.
 DR P-PSDB; AAU20977.
 XX New polynucleotides and polypeptides for diagnosing, treating,
 PT preventing or prognosing e.g. diseases or disorders of the nervous,
 PT musculoskeletal, excretory, gastrointestinal, reproductive, and
 PT respiratory systems
 XX Claim 1; SEQ ID NO 321; 642pp; English.
 XX The invention relates to novel nucleic acids encoding novel human foetal
 CC antigens. The nucleic acids and proteins are used to prevent, treat (e.g.
 CC by gene therapy) or ameliorate a medical condition in e.g. humans, mice,
 CC rabbits, goats, horses, cats, dogs, chickens or sheep. They
 CC are also used in diagnosing a pathological condition or susceptibility
 CC to a pathological condition. The antibodies to the antigens can also
 CC be used in alleviating symptoms associated with the disorders and in
 CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked
 CC immunosorbent assays (ELISA). Disorders which are diagnosed or treated
 CC include autoimmune diseases e.g. rheumatoid arthritis,
 CC hyperproliferative disorders e.g. neoplasms of the breast or liver,
 CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
 CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.
 CC Alzheimer's disease, infections caused by bacteria, viruses and fungi
 CC and ocular disorders e.g. corneal infection. The polypeptides can also
 CC be used to aid wound healing and epithelial cell proliferation, to
 CC prevent skin aging due to sunburn, to maintain organs before
 CC transplantation, for supporting cell culture of primary tissues, to
 CC regenerate tissues and in chemotaxis. The polypeptides can also be used
 CC as a food additive or preservative to increase or decrease storage
 CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
 CC minerals, cofactors and other nutritional components. Numerous
 CC examples of diseases and disorders treated by the nucleic acids and
 CC proteins are given in the specification. The present sequence

Query Match 2.0%; Score 29; DB 22; Length 428;
 Best Local Similarity 100.0%; Pred. No. 0.0007;
 Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 825 GAGACCTGGCTGCGCCCACTACAT 853
 |||||
 Db 1 GAGACCTGGCTGCGCCCACTACAT 29

RESULT 11
 ABT03700/C
 ID ABT03700 standard; DNA; 26 BP.
 AC ABT03700;
 XX
 XX 13-SEP-2002 (first entry)
 DT
 XX Human Neurogenin-3 gene PCR primer SEQ ID NO: 221.
 DE
 XX Human; cancer; neoplastic disease; tumour specific marker; cytostatic;
 KW transcription factor; PCR; primer; ss.
 KW
 XX Homo sapiens.
 OS
 XX W0200240716-A2.
 PN
 XX 23-MAY-2002.
 PD
 XX 13-NOV-2001; 2001WO-US43461.
 PF
 XX 16-NOV-2000; 2000US-249508P.
 PR
 XX (CEMT-) CEMINES LLC.
 PA
 XX Palm K;
 PI

XX WPI; 2002-537346/57.
 XX
 PT Determining the presence of neoplastic molecular markers, by
 PT identifying the presence of markers in host test sample using array of
 PT neoplastic molecular marker specific reagents and analyzing the array
 PT of the reagents
 XX Example 1; Page 17; 41pp; English.
 XX
 CC The present invention relates to a method for determining the presence of
 CC neoplastic molecular markers in a host, involving the use of neoplastic
 CC molecular marker specific reagents to detect such markers and analysing
 CC the array of reagents, allowing the identification of the neoplastic
 CC disease present. This can be used to determine the best treatment for
 CC cancers, in particular neural cell, lung and prostate tumours. The
 CC present sequence is a PCR primer useful for detecting the coding
 CC sequences of markers of the invention.
 XX
 SQ Sequence 26 BP; 3 A; 10 C; 5 G; 8 T; 0 other;
 Query Match 1.8%; Score 26; DB 24; Length 26;
 Best Local Similarity 100.0%; Pred. No. 0.024;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 673 AGAGCGAGTTGGCACTGAGCAAGCAG 698
 |||||
 Db 26 AGAGCGAGTTGGCACTGAGCAAGCAG 1
 RESULT 12
 AAV42515
 ID AAV42515 standard; DNA; 25 BP.
 XX
 AC AAV42515;
 XX
 DT 05-OCT-1998 (first entry)
 XX
 DE PCR primer used to isolate part of the RELAX protein coding region.
 KW Basic helix-loop-helix; BHLH; RELAX; Rat Embryonic Longitudinal Axis;
 KW control; gene expression; transcriptional activator; targeting;
 KW protein expression; central nervous system; CNS; treatment;
 KW nervous system disorder; CIG235; PCR primer; ss.
 XX
 OS Synthetic.
 OS Rattus sp.
 XX
 PN W09827206-A2.
 XX
 XX 25-JUN-1998.
 PD
 XX 19-DEC-1997; 97WO-FR02368.
 PF
 XX 19-DEC-1996; 96FR-0015651.
 PR
 XX (RHON) RHONE-POULENC RORER SA.
 PA
 XX Mallet J, Ravassard P, Icard-Liepkalns C;
 PI
 XX WPI; 1998-362775/31.
 DR
 XX Basic helix-loop-helix polypeptide and related nucleic acid - with
 PT transcriptional activity, for targeting expression of genes to
 PT central nervous system and treatment of nervous disease
 PT
 XX Example 2; Page 12; 28pp; French.
 PS
 XX PCR primers AAV42515-16 are used to isolate part of the DNA encoding
 CC a basic helix-loop-helix (BHLH) type protein, designated RELAX (Rat
 CC Embryonic Longitudinal Axis) protein. The PCR product is termed CIG235.
 CC The protein is used to control and participate in gene expression,
 CC by acting as transcriptional activator, strictly dependent on the

CC presence of an intact E box (CANNTG), particularly for targeting
 CC expression of proteins to the central nervous system (CNS). The
 CC nucleic acid sequence can be used to treat nervous system disorders,
 CC and antisense sequences can be used to control mRNA transcription.

XX Sequence 25 BP; 5 A; 9 C; 6 G; 5 T; 0 other;
 SQ Query Match 1.7%; Score 25; DB 19; Length 25;
 Best Local Similarity 100.0%; Pred. No. 0.073;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 750 AACCTTAACCTCGCGTGGATGCGC 774

DB 1 AACCTTAACCTCGCGTGGATGCGC 25

RESULT 13

AAV27046

ID AAV27046 standard; cDNA; 738 BP.

XX AC AAV27046;

XX AC AAV27046;

XX 17-SEP-1998 (first entry)

XX Mouse neurogenin 1 gene.

XX ss; Mouse; neurogenin; expression vector; recombinant protein;

XX antibody; neurogenesis.

XX Mus sp.

XX Key

XX CDS

XX Location/Qualifiers

XX 1..735

XX /tag- a

XX /product- "Mouse neurogenin 1"

XX WO9813491-A2.

XX 02-APR-1998.

XX 24-SEP-1997;

XX 97WO-US17048.

XX 17-SEP-1997;

XX 97US-0932411.

XX 27-SEP-1996;

XX 96US-0722570.

XX 12-NOV-1996;

XX 96US-0030864.

XX 19-DEC-1996;

XX 96US-0772009.

XX (CALY) CALIFORNIA INST OF TECHNOLOGY.

XX Anderson DJ, Ma Q, Sommer L;

XX WPI; 1998-230702/20.

XX P-PSDB; AAW54944.

XX Mouse neurogenin1s, useful in neurogenesis - and recombinant nucleic

XX acids and proteins derived from rat and Xenopus

XX Claim 5; Fig 4; 106pp; English.

XX The mouse neurogenin 1 is one of several neurogenin proteins discussed in

XX the present invention. The neurogenin nucleic acids can be expressed in

XX a host cell, transformed using an expression vector, to produce

XX recombinant proteins. The proteins and the antibodies raised against

XX the proteins are useful in the study of neurogenesis.

XX Sequence 738 BP; 121 A; 283 C; 205 G; 129 T; 0 other;

XX Query Match 1.6%; Score 23; DB 19; Length 738;

XX Best Local Similarity 100.0%; Pred. No. 0.59;

XX Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX 732 GAGCGCAACCGCATGCACACCT 754

XXXXXXXXXXXXXXXXXXXXXXXXXXXX

DB 304 GAGCGCAACCGCATGCACACCT 326

RESULT 14

AAZ51977

ID AAZ51977 standard; DNA; 738 BP.

XX AC AAZ51977;

XX AC AAZ51977;

XX 04-JUL-2000 (first entry)

XX Murine neurogenin-1 (NGN1) nucleic acid sequence.

XX Neurogenin-1; NGN-1; non-neuronal cell; NNC; neurogenesis;

XX Phox2a protein; neuronal subtype-specific marker; growth factor;

XX neural differentiation; transplantation; neuronal dysfunction;

XX optical nerve damage; auditory nerve damage; neurodegenerative disorder;

XX neuroprotective; neurotropic; anticonvulsant; antiParkinsonian; vulnary;

XX cerebroprotective; immunosuppressant; antiinfectious; ss.

XX Mus sp.

XX Key

XX CDS

XX Location/Qualifiers

XX 1..735

XX /tag- a

XX /product- "Murine neurogenin-1 protein"

XX WO200009676-A2.

XX 24-FEB-2000.

XX 13-AUG-1999;

XX 99WO-US18525.

XX 14-AUG-1998;

XX 98US-0096630.

XX (CALY) CALIFORNIA INST OF TECHNOLOGY.

XX Anderson DJ, Lo L;

XX WPI; 2000-256250/22.

XX P-PSDB; AAY70566.

XX Inducing non-neuronal cells to differentiate into neurons and for

XX non-neuronal cells to express a neuronal subtype-specific marker,

XX comprising contacting the non-neuronal cells with a vector containing

XX neurogenin nucleic acid -

XX Claim 1; Fig 1C; 76pp; English.

XX The patent discloses a method for inducing non-neuronal cells (NNC) to

XX differentiate into neurons and for NNCs to express a neuronal subtype

XX -specific marker. Transformed host cells are used as sources of neuronal

XX and other growth factors; in culture for screening compounds that

XX modulate neural differentiation or as sources of recombinantly produced

XX neurogenins and Phox2a proteins for use in transplantation. The cells

XX also have a variety of in vivo uses, e.g. for transplantation at sites of

XX neuronal dysfunction e.g. patients with hearing or vision loss due to

XX optical or auditory nerve damage, brain or spinal cord injuries, and

XX neurodegenerative disorders e.g. Alzheimer's disease. The present

XX sequence encodes murine neurogenin-1 (NGN-1), a transcription factor.

XX NNCs differentiate into neurons through the recombinant expression of a

XX transcription factor that induces a core program of neurogenesis. Forced

XX expression of murine NGN1 can elicit expression of at least some neuronal

XX phenotypic markers even in NNCs. This can be used in autografting.

XX Sequence 738 BP; 121 A; 283 C; 205 G; 129 T; 0 other;

XX Query Match 1.6%; Score 23; DB 21; Length 738;

XX Best Local Similarity 100.0%; Pred. No. 0.59;

XX Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX 732 GAGCGCAACCGCATGCACACCT 754

XXXXXXXXXXXXXXXXXXXXXXXXXXXX

Db 304 GAGCGCAACCGCATGCACACCT 326

Search completed: April 8, 2003, 19:05:41
Job time : 407 secs

RESULT 15

AAF27264
ID AAF27264 standard; CDNA; 790 BP.
XX AC AAF27264;
XX DT
XX XX
XX 24-APR-2001 (first entry)
XX DE
XX DE Chicken atonal homologue ngn2/ath4a cDNA, SEQ ID NO:20.
XX KW Atonal; homologue; orthologue; atonal-associated protein; deafness;
KW hearing impairment; vestibular effect; balance disorder; osteoarthritis;
KW cellular proliferation; cerebellar granule neuron; gene therapy;
KW mechanoreceptive cell growth; auditory; osteopathic; cytostatic;
KW transgenic animal; ss.
XX OS
XX OS Gallus gallus.
XX XX
XX PN WO200073764-A2.
XX XX
XX PD 07-DEC-2000.
XX PF 01-JUN-2000; 2000WO-US15410.
XX XX
XX PR 01-JUN-1999; 99US-0137060.
XX PR 19-JAN-2000; 2000US-0176993.
XX XX
XX PA (BAYU) BAYLOR COLLEGE MEDICINE.
XX XX
XX PI Zoghbi HY, Bellen H, Birmingham N, Hassan B, Ben-Arie N;
XX P-PSDB; AAB60357.
XX DR
XX DR
XX PT Therapeutic use of atonal-associated nucleic acids or amino acids, or
PT any of its homologs or orthologs, for the treatment of e.g. deafness,
PT osteoarthritis and abnormal cell proliferation -
XX XX
XX PS Disclosure; Page -: 142pp; English.
XX XX
XX CC The invention relates to the use of atonal-associated nucleic acid or
CC amino acid sequence, or any of its homologues or orthologues as
CC therapeutic agents for the treatment of deafness, partial hearing loss,
CC vestibular effects due to damage or loss of inner hair cells,
CC osteoarthritis and abnormal cell proliferation. The invention also
CC encompasses methods of screening for compounds which affect the
CC expression of an atonal-associated nucleic acid sequence in an animal,
CC and a transgenic animal in which an allele of a native atonal-associated
CC gene is replaced by a heterologous nucleic acid sequence, thus
CC inactivating the atonal-associated allele. The nucleic acids or proteins
CC may be used in a method of treating an animal for hearing impairment,
CC joint disease, balance disorders, abnormal cell proliferation, or other
CC disease related to loss of a functional atonal-associated nucleic acid or
CC protein. They may particularly be used to treat an animal with a
CC deficiency in cerebellar granule neurons or their precursors, and may
CC also be used in promoting mechanoreceptive cell growth and generating
CC hair cells. The present sequence represents an atonal-associated nucleic
CC acid sequence referred to in the invention.
CC Note: The present sequence is not shown in the specification, but
CC was obtained from GenBank.
XX XX

XX SQ Sequence 790 BP; 91 A; 351 C; 283 G; 65 T; 0 other;

Query Match 1.6%; Score 23; DB 22; Length 790;

Best Local Similarity 100.0%; Pred. No. 0.59;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 732 GAGCGCAACCGCATGCACACCT 754

|||||

Db 374 GAGCGCAACCGCATGCACACCT 396

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 8, 2003, 17:56:31 ; Search time 165 seconds
(without alignments)
7761.606 Million cell updates/sec

Title: US-09-595-947C-1

Perfect score: 1460

Sequence: 1 gcaggtacgagagagcag.....agagtacacacacacagcgt 1460

Scoring table: OLIGO.NUC

Gapop 60.0 , Gapext 60.0

Searched: 593429 seqs, 438583890 residues

Word size : 8

Total number of hits satisfying chosen parameters: 919487

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

Published_Applications_NA.*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/2/pubpna/US05_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
- 10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
- 11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
- 12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq.*
- 13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
- 14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	92	6.3	645	9	US-10-004-7117-4
2	92	6.3	861	9	US-10-004-7117-24
3	92	6.3	1861	10	US-09-817-360-3
4	32	2.2	5340	10	US-09-817-360-1
5	23	1.6	738	7	US-08-722-570-13
6	23	1.6	790	9	US-10-004-7117-20
7	23	1.6	1385	9	US-10-004-7117-30
8	23	1.6	1412	9	US-10-004-7117-6
9	23	1.6	1412	9	US-10-004-7117-37
10	20	1.4	352	10	US-09-728-445-108
11	19	1.3	500	10	US-09-783-590-2157
12	19	1.3	1263	9	US-09-938-842A-1036
13	18	1.2	4797	10	US-09-751-797-25
14	18	1.2	20272	10	US-09-908-711-145
15	18	1.2	25619	10	US-09-908-711-143
16	18	1.2	25619	10	US-09-764-898-302
17	18	1.2	31728	9	US-10-114-170-64
18	18	1.2	402850	9	US-09-844-653-5
19	17	1.2	31	10	US-09-817-360-17

20	17	1.2	232	10	US-09-960-352-14228	Sequence 14228, A
21	17	1.2	285	10	US-09-960-352-6887	Sequence 6887, Ap
22	17	1.2	285	10	US-09-960-352-8778	Sequence 8778, Ap
c 23	17	1.2	307	10	US-09-783-590-12149	Sequence 12149, A
24	17	1.2	454	9	US-10-040-739-1094	Sequence 1094, Ap
c 25	17	1.2	454	10	US-09-864-761-257	Sequence 257, App
c 26	17	1.2	469	10	US-09-864-761-5374	Sequence 5374, Ap
c 27	17	1.2	471	10	US-09-864-761-11	Sequence 11, Appl
c 28	17	1.2	473	10	US-09-864-761-22148	Sequence 22148, A
c 29	17	1.2	474	10	US-09-864-761-353	Sequence 353, App
c 30	17	1.2	474	10	US-09-864-761-10932	Sequence 10932, A
c 31	17	1.2	487	10	US-09-864-761-3320	Sequence 3320, Ap
32	17	1.2	495	10	US-09-864-761-5673	Sequence 5673, Ap
c 33	17	1.2	499	10	US-09-783-590-4421	Sequence 4421, Ap
c 34	17	1.2	594	10	US-09-864-761-7938	Sequence 7938, Ap
c 35	17	1.2	756	10	US-09-910-943-147	Sequence 147, App
36	17	1.2	1074	9	US-10-004-7117-18	Sequence 18, Appl
c 37	17	1.2	1238	9	US-09-991-496-75	Sequence 75, Appl
c 38	17	1.2	1238	10	US-09-874-923-75	Sequence 75, Appl
c 39	17	1.2	1362	9	US-09-970-966-208	Sequence 208, App
c 40	17	1.2	1362	10	US-09-825-294-208	Sequence 208, App
41	17	1.2	1422	10	US-09-815-242-4334	Sequence 4334, Ap
42	17	1.2	1425	10	US-09-815-242-8558	Sequence 8558, Ap
43	17	1.2	1669	10	US-09-880-107-2159	Sequence 2159, Ap
44	17	1.2	1747	9	US-09-764-868-1334	Sequence 1334, Ap
45	17	1.2	1752	9	US-09-764-868-137	Sequence 137, App

ALIGNMENTS

RESULT 1

US-10-004-7117-4
; Sequence 4, Application US/10004717
; Publication No. US20020192665A1
; GENERAL INFORMATION:
; APPLICANT: ZOGHBI, HUDA Y.
; APPLICANT: YANG, QI
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN ATONAL ASSOCIATED SEQUENCE FOR DEAFNESS.
; TITLE OF INVENTION: OSTEOARTHRITIS, AND ABNORMAL CELL PROLIFERATION
; FILE REFERENCE: P01899054
; CURRENT APPLICATION NUMBER: US/10/004,717
; CURRENT FILING DATE: 2002-08-16
; PRIOR APPLICATION NUMBER: 09/585,645
; PRIOR FILING DATE: 2000-06-01
; PRIOR APPLICATION NUMBER: 60/176,993
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: 60/137,060
; PRIOR FILING DATE: 1999-06-01
; NUMBER OF SEQ ID NOS: 69
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 645
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-004-7117-4

Query Match 6.3%; Score 92; DB 9; Length 645;

Best Local Similarity 100.0%; Pred. No. 5.5e-39;

Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 762 GGCGTGGTGGCGTGGCGGTGTCCTGCCACCTTCCCGATGACGCCAACCTTACCAAG 821

Db 304 GGCGTGGTGGCGTGGCGGTGTCCTGCCACCTTCCCGATGACGCCAACCTTACCAAG 363

QY 822 ATCGAGACCTCGGCTTCGCCCACTACAT 853

Db 364 ATCGAGACCTCGGCTTCGCCCACTACAT 395

RESULT 2

US-10-004-7117-24

; Sequence 24, Application US/10004717
; Publication No. US20020192665A1
; GENERAL INFORMATION:
; APPLICANT: YANG, QI
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN
; TITLE OF INVENTION: ATONAL ASSOCIATED SEQUENCE FOR DEAFNESS,
; TITLE OF INVENTION: OSTEOARTHRITIS, AND ABNORMAL CELL PROLIFERATION
; FILE REFERENCE: P01899US4
; CURRENT APPLICATION NUMBER: US/10/004,717
; PRIOR FILING DATE: 2002-08-16
; PRIOR APPLICATION NUMBER: 09/585,645
; PRIOR FILING DATE: 2000-06-01
; PRIOR APPLICATION NUMBER: 60/176,993
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: 60/137,060
; PRIOR FILING DATE: 1999-06-01
; NUMBER OF SEQ ID NOS: 69
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 24
; LENGTH: 861
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-004-717-24

Query Match 6.3%; Score 92; DB 9; Length 861;
Best Local Similarity 100.0%; Pred. No. 5.4e-39;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 762 GCGCTGGATCGCTGCGCGGTGCTCCGCCACCTTCCCGGATGACGCCAACTTACAAAG 821
DB 463 GCGCTGGATCGCTGCGCGGTGCTCCGCCACCTTCCCGGATGACGCCAACTTACAAAG 522
QY 822 ATCGAGACCTGCGCTTCGCCCACTACAT 853
DB 523 ATCGAGACCTGCGCTTCGCCCACTACAT 554

RESULT 3
US-09-817-360-3
; Sequence 3, Application US/09817360
; Patent No. US20020015696A1
; GENERAL INFORMATION:
; APPLICANT: German, Michael S.
; APPLICANT: Lin, Joseph
; TITLE OF INVENTION: PRODUCTION OF PANCREATIC ISLET CELLS
; TITLE OF INVENTION: AND DELIVERY OF INSULIN
; FILE REFERENCE: UCSF-129CIP
; CURRENT APPLICATION NUMBER: US/09/817,360
; CURRENT FILING DATE: 2001-03-20
; PRIOR FILING DATE: 2000-03-24
; PRIOR APPLICATION NUMBER: 09/535,145
; PRIOR FILING DATE: 2000-03-24
; PRIOR APPLICATION NUMBER: 60/128,180
; PRIOR FILING DATE: 1999-04-06
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1861
; TYPE: DNA
; ORGANISM: Mus musculus
US-09-817-360-3

Query Match 6.3%; Score 92; DB 10; Length 1861;
Best Local Similarity 100.0%; Pred. No. 5e-39;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 762 GCGCTGGATCGCTGCGCGGTGCTCCGCCACCTTCCCGGATGACGCCAACTTACAAAG 821
DB 1396 GCGCTGGATCGCTGCGCGGTGCTCCGCCACCTTCCCGGATGACGCCAACTTACAAAG 1455
QY 822 ATCGAGACCTGCGCTTCGCCCACTACAT 853
DB 1456 ATCGAGACCTGCGCTTCGCCCACTACAT 1487

RESULT 4
US-09-817-360-1
; Sequence 1, Application US/09817360
; Patent No. US20020015696A1
; GENERAL INFORMATION:
; APPLICANT: German, Michael S.
; APPLICANT: Lin, Joseph
; TITLE OF INVENTION: PRODUCTION OF PANCREATIC ISLET CELLS
; TITLE OF INVENTION: AND DELIVERY OF INSULIN
; FILE REFERENCE: UCSF-129CIP
; CURRENT APPLICATION NUMBER: US/09/817,360
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: 09/535,145
; PRIOR FILING DATE: 2000-03-24
; PRIOR APPLICATION NUMBER: 60/128,180
; PRIOR FILING DATE: 1999-04-06
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 5340
; TYPE: DNA
; ORGANISM: Homo Sapiens
US-09-817-360-1

Query Match 2.2%; Score 32; DB 10; Length 5340;
Best Local Similarity 100.0%; Pred. No. 4.9e-07;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 672 AAGAGCGAGTGGCACTGACGACGACGACG 703
DB 3235 AAGAGCGAGTGGCACTGACGACGACGACG 3266

RESULT 5
US-08-722-570-13
; Sequence 13, Application US/08722570
; Publication No. US20030044887A1
; GENERAL INFORMATION:
; APPLICANT: Anderson, David J.
; APPLICANT: Ma, Qiufu
; TITLE OF INVENTION: NEUROGENIN
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/722,570
; FILING DATE: 27-SEP-1996
; CLASSIFICATION: 5365
; ATTORNEY/AGENT INFORMATION:
; NAME: Silva, Robin M.
; REGISTRATION NUMBER: 38,304
; REFERENCE/DOCKET NUMBER: A-63902/RFT/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 738 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown

```
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA
US-08-722-570-13

Query Match      1.6%; Score 23; DB 7; Length 738;
Best Local Similarity 100.0%; Pred. No. 0.038;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 732 GAGCGCAACCGCATGCACAACCT 754
|||||
Db 304 GAGCGCAACCGCATGCACAACCT 326

RESULT 6
US-10-004-717-20
; Sequence 20, Application US/10004717
; Publication No. US20020192665A1
; GENERAL INFORMATION:
; APPLICANT: ZOGHBI, HUDA Y.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN
; TITLE OF INVENTION: ATONAL ASSOCIATED SEQUENCE FOR DEAFNESS,
; TITLE OF INVENTION: OSTEOARTHRITIS, AND ABNORMAL CELL PROLIFERATION
; FILE REFERENCE: P01899US4
; CURRENT APPLICATION NUMBER: US/10/004,717
; CURRENT FILING DATE: 2002-08-16
; PRIOR APPLICATION NUMBER: 09/585,645
; PRIOR FILING DATE: 2000-06-01
; PRIOR APPLICATION NUMBER: 60/176,993
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: 60/137,060
; PRIOR FILING DATE: 1999-06-01
; NUMBER OF SEQ ID NOS: 69
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 20
; LENGTH: 790
; TYPE: DNA
; ORGANISM: chicken
US-10-004-717-20

Query Match      1.6%; Score 23; DB 9; Length 790;
Best Local Similarity 100.0%; Pred. No. 0.038;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 732 GAGCGCAACCGCATGCACAACCT 754
|||||
Db 374 GAGCGCAACCGCATGCACAACCT 396

RESULT 7
US-10-004-717-30
; Sequence 30, Application US/10004717
; Publication No. US20020192665A1
; GENERAL INFORMATION:
; APPLICANT: ZOGHBI, HUDA Y.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN
; TITLE OF INVENTION: ATONAL ASSOCIATED SEQUENCE FOR DEAFNESS,
; TITLE OF INVENTION: OSTEOARTHRITIS, AND ABNORMAL CELL PROLIFERATION
; FILE REFERENCE: P01899US4
; CURRENT APPLICATION NUMBER: US/10/004,717
; CURRENT FILING DATE: 2002-08-16
; PRIOR APPLICATION NUMBER: 09/585,645
; PRIOR FILING DATE: 2000-06-01
; PRIOR APPLICATION NUMBER: 60/176,993
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: 60/137,060
; PRIOR FILING DATE: 1999-06-01
; NUMBER OF SEQ ID NOS: 69
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 30
; LENGTH: 1385
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-004-717-30

Query Match      1.6%; Score 23; DB 9; Length 1385;
Best Local Similarity 100.0%; Pred. No. 0.036;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 732 GAGCGCAACCGCATGCACAACCT 754
|||||
Db 742 GAGCGCAACCGCATGCACAACCT 764

RESULT 8
US-10-004-717-6
; Sequence 6, Application US/10004717
; Publication No. US20020192665A1
; GENERAL INFORMATION:
; APPLICANT: ZOGHBI, HUDA Y.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN
; TITLE OF INVENTION: ATONAL ASSOCIATED SEQUENCE FOR DEAFNESS,
; TITLE OF INVENTION: OSTEOARTHRITIS, AND ABNORMAL CELL PROLIFERATION
; FILE REFERENCE: P01899US4
; CURRENT APPLICATION NUMBER: US/10/004,717
; CURRENT FILING DATE: 2002-08-16
; PRIOR APPLICATION NUMBER: 09/585,645
; PRIOR FILING DATE: 2000-06-01
; PRIOR APPLICATION NUMBER: 60/176,993
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: 60/137,060
; PRIOR FILING DATE: 1999-06-01
; NUMBER OF SEQ ID NOS: 69
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 1412
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-004-717-6

Query Match      1.6%; Score 23; DB 9; Length 1412;
Best Local Similarity 100.0%; Pred. No. 0.036;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 732 GAGCGCAACCGCATGCACAACCT 754
|||||
Db 445 GAGCGCAACCGCATGCACAACCT 467

RESULT 9
US-10-004-717-37
; Sequence 37, Application US/10004717
; Publication No. US20020192665A1
; GENERAL INFORMATION:
; APPLICANT: ZOGHBI, HUDA Y.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN
; TITLE OF INVENTION: ATONAL ASSOCIATED SEQUENCE FOR DEAFNESS,
; TITLE OF INVENTION: OSTEOARTHRITIS, AND ABNORMAL CELL PROLIFERATION
; FILE REFERENCE: P01899US4
; CURRENT APPLICATION NUMBER: US/10/004,717
; CURRENT FILING DATE: 2002-08-16
; PRIOR APPLICATION NUMBER: 09/585,645
; PRIOR FILING DATE: 2000-06-01
; PRIOR APPLICATION NUMBER: 60/176,993
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: 60/137,060
; PRIOR FILING DATE: 1999-06-01
; NUMBER OF SEQ ID NOS: 69
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 37
; LENGTH: 1412
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-004-717-37
```

US-10-004-717-37

Query Match 1.6%; Score 23; DB 9; Length 1412;
Best Local Similarity 100.0%; Pred. No. 0.036;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 732 GAGCGCAACCGCATGCACAACT 754
|||||
DB 445 GAGCGCAACCGCATGCACAACT 467

RESULT 10

US-09-728-445-108/c
; Sequence 108, Application US/09728445
; Patent No. US20020102543A1
; GENERAL INFORMATION:
; APPLICANT: Friedrich, Glenn
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: NO. US20020102543A1el Mutated Mammalian Cells and
; FILE REFERENCE: LEX-0102-USA
; CURRENT APPLICATION NUMBER: US/09/728,445
; CURRENT FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/168,358
; PRIOR FILING DATE: 1999-12-01
; NUMBER OF SEQ ID NOS: 891
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 108
; LENGTH: 352
; TYPE: DNA
; ORGANISM: Mus musculus
US-09-728-445-108

Query Match 1.4%; Score 20; DB 10; Length 352;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1010 CACAGCCTCATTTGGAGGACT 1029
|||||
DB 209 CACAGCCTCATTTGGAGGACT 190

RESULT 11

US-09-783-590-2157/c
; Sequence 2157, Application US/09783590
; Patent No. US20020110850A1
; GENERAL INFORMATION:
; APPLICANT: Dillon, Patrick J.
; APPLICANT: Haseltine, William A.
; APPLICANT: Li, Haodong
; APPLICANT: Rosen, Craig A.
; APPLICANT: Ruben, Steven M.
; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2
; FILE REFERENCE: PO-16.2C1
; CURRENT APPLICATION NUMBER: US/09/783,590
; CURRENT FILING DATE: 2000-02-15
; PRIOR APPLICATION NUMBER: 08/420,856
; PRIOR FILING DATE: 1995-04-12
; PRIOR APPLICATION NUMBER: 08/346,731
; PRIOR FILING DATE: 1994-11-21
; NUMBER OF SEQ ID NOS: 12485
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2157
; LENGTH: 500
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (84)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (207)

; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (215)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (324)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (354)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (360)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (368)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (369)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (379)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (393)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (411)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (418)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (439)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (440)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (441)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (451)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-783-590-2157

Query Match 1.3%; Score 19; DB 10; Length 500;
Best Local Similarity 100.0%; Pred. No. 5.4;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1319 CCACCTCAAACTCCCGCTC 1337
|||||
DB 23 CCACCTCAAACTCCCGCTC 5

RESULT 12

US-09-938-842A-1036/c
; Sequence 1036, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Krieps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAININ
; FILE REFERENCE: SAME, AND METHODS OF USE
; FILE REFERENCE: SCRIPI300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111

; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 1036
; LENGTH: 1263
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-1036

Query Match 1.3%; Score 19; DB 9; Length 1263;
Best Local Similarity 100.0%; Pred. No. 4.9;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 658 GCGCGAACAGGCCCAAGAG 676
|||||
DB 599 GCGCGAACAGGCCCAAGAG 581

RESULT 13

US-09-751-797-25
; Sequence 25, Application US/09751797
; Patent No. US20010024652A1

; GENERAL INFORMATION:
; APPLICANT: Dumoutier, Laure
; APPLICANT: Louhed, Jamila
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules which Encode T Cell Inducible Factors
; TITLE OF INVENTION: (TIFS) The Proteins Encoded, and Uses Thereof
; FILE REFERENCE: LUD 5543.2
; CURRENT APPLICATION NUMBER: US/09/751,797
; CURRENT FILING DATE: 2000-12-29
; PRIOR APPLICATION NUMBER: 09/419,568
; PRIOR FILING DATE: 1999-10-18
; PRIOR APPLICATION NUMBER: US09/178,973
; PRIOR FILING DATE: 1998-10-26
; NUMBER OF SEQ ID NOS: 29
; SEQ ID NO 25
; LENGTH: 4797
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
US-09-751-797-25

Query Match 1.2%; Score 18; DB 10; Length 4797;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1152 GTCAGAGCTGTCTGAAAT 1169
|||||
DB 3460 GTCAGAGCTGTCTGAAAT 3477

RESULT 14

US-09-908-711-145/c
; Sequence 145, Application US/09908711
; Patent No. US20020045230A1

; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA128
; CURRENT APPLICATION NUMBER: US/09/908,711
; CURRENT FILING DATE: 2001-07-20
; PRIOR APPLICATION NUMBER: US01/01360
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,867
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01344
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,892
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01345
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,888
; PRIOR FILING DATE: 2001-01-17

; PRIOR APPLICATION NUMBER: US01/01329
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,905
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01354
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,891
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01339
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,869
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01340
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,874
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01334
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,898
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01320
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,853
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01349
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,902
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01239
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,870
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01348
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,882
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01347
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,896
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01307
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,864
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01341
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,856
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01336
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,868
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01312
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: 60/209,467
; PRIOR FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 167
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 145
; LENGTH: 20272
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-908-711-145

Query Match 1.2%; Score 18; DB 10; Length 20272;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1153 TCAGAGCTGTCTGAAATG 1170

Db 6642 TCAGAGCTGTCTGAAATG 6625
|||||

RESULT 15

US-09-908-711-143/c
; Sequence 143, Application US/09908711
; Patent No. US20020045230A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA128
; CURRENT APPLICATION NUMBER: US/09/908,711
; CURRENT FILING DATE: 2001-07-20
; PRIOR APPLICATION NUMBER: US01/01360
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,867
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01344
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,892
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01345
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,888
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01329
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,905
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01354
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,891
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01339
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,869
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01340
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,874
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01334
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,898
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01320
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,853
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01349
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,902
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01239
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,870
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01348
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,882
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01347
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,896
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01307
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,864
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01341
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,856

; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01336
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,868
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01312
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: 60/209,467
; PRIOR FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 167
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 143
; LENGTH: 25619
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-908-711-143

Query Match 1.2%; Score 18; DB 10; Length 25619;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Caps 0;

QY 1153 TCAGAGCTGTCTGAAATG 1170
|||||
Db 1048 TCAGAGCTGTCTGAAATG 1031

Search completed: April 8, 2003, 21:14:29
Job time : 291 secs

Query Match	1.4%; Score 21; DB 9;	Length 953;
Best Local Similarity	100.0%;	Pred. No. 19;
Matches 21; Conservative 0;	Mismatches 0;	Indels 0; Gaps 0;
Qy 729	CGGGAGCGCAACCGCATGCAC	749
Dd 334	CGGGAGCGCAACCGCATGCAC	354
RESULT 14		
BQ686909		
LOCUS	BQ686909	962 bp mRNA linear EST 15-JUL-2002
DEFINITION	AGENCOCURT-8340711 NIH_MGC_110 Homo sapiens cDNA clone IMAGE:6249314	
	5'..	mRNA sequence.

ACCESSION	BQ686909	
VERSION	BQ686909.1	GI:21812225
KEYWORDS	EST.	
SOURCE	human.	
ORGANISM	Homo sapiens	
	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;	
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
REFERENCE	1 (bases 1 to 962)	

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabs-r@mail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LCM2390 row: j column: 03
 High quality sequence start: 3
 High quality sequence stop: 587.

```

Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC library."
BASE COUNT      236 a      273 c      239 g      194 t
ORIGIN

Query Match      1.4%; Score 21; DB 14; Length 962;
Best Local Similarity 100.0%; Pred. NO. 19;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 429 ACCCCCTCCATTTTTCCTCC 449
      |||||||TTTTTTTTTTTT|
Db 811 ACCCCCTCCATTTTTCCTCC 831

RESULT 15
LOCUS      BM563662
DEFINITION AGENCOURT_6589536 NIH_MGC_98 Homo sapiens cDNA clone IMAGE:5441798
            5', mRNA sequence.
ACCESSION  BM563662
VERSION     BM563662.1 GI:18810777

```


KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1022)
AUTHORS NIH-MGC <http://mgs.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLCM1916 row: k column: 15
High quality sequence stop: 654.
FEATURES
Location/Qualifiers
1..1022
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5441798"
/clone_lib="NIH_MGC_98"
/tissue_type="astrocytoma grade IV, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: brain; Vector: pORF7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into ECORI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Library constructed by Ling Hong
in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."
BASE COUNT 177 a 324 c 319 g 202 t
ORIGIN
Query Match 1.4%; Score 21; DB 13; Length 1022;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1112 AACAGGCCCTGGCGGTGGGC 1132
|||||
DB 573 AACAGGCCCTGGCGGTGGGC 593
Search completed: April 8, 2003, 21:09:45
Job time : 2284 secs

pred. No. is the number of results predicted by chance to have a

QY 301 GCAGCTCTCTGTTCTTTTGAGCCCGGAGTAACTAGGTAACATTTAGGAACCTCCAAAGG 360

```

Db 301 GCAGCTCTCTGTTCTTTTGGCCGCGAGTAAGTAACTTAGGAACCTCCAAAGG 360
QY 361 TAGAAGAGGGAGTGGTGGGTACTCTAGTCCCGCGTGGAGTACCTCTAAGTCAGAG 420
Db 361 TAGAAGAGGGAGTGGTGGGTACTCTAGTCCCGCGTGGAGTACCTCTAAGTCAGAG 420
QY 421 ACTGTCCACACCCCTCTCCATTTTCCCAACCTCAGATGGCGCTCATCCCTTGGATG 480
Db 421 ACTGTCCACACCCCTCTCCATTTTCCCAACCTCAGATGGCGCTCATCCCTTGGATG 480
QY 481 CGCCCAACATCCAAGTGTCCCAAGAGACCCAGAACCTTTCCCGAGGCTCGGACCAGG 540
Db 481 CGCCCAACATCCAAGTGTCCCAAGAGACCCAGAACCTTTCCCGAGGCTCGGACCAGG 540
QY 541 AAGTGTCTCAGTTCCCAATTCACCCCACTAGCCCACTCTCGTACCGAGGACTGCTCCG 600
Db 541 AAGTGTCTCAGTTCCCAATTCACCCCACTAGCCCACTCTCGTACCGAGGACTGCTCCG 600
QY 601 AAGCAGAAGCAGTGTACTCCGAGGACATCGAGGAAGCTCCGTGCGCGCGGAGGCG 660
Db 601 AAGCAGAAGCAGTGTACTCCGAGGACATCGAGGAAGCTCCGTGCGCGCGGAGGCG 660
QY 661 GCACAGGCGCCAAAGAGGAGTTGGCACTAGCAAGCAGCAGAGCCCGGCCAAGAAGG 720
Db 661 GCACAGGCGCCAAAGAGGAGTTGGCACTAGCAAGCAGCAGAGCCCGGCCAAGAAGG 720
QY 721 CCAACGACCGGAGCGCAACCGCATGCAACCTTAACCTCCGCTGGATGCGCTCGCG 780
Db 721 CCAACGACCGGAGCGCAACCGCATGCAACCTTAACCTCCGCTGGATGCGCTCGCG 780
QY 781 GTGTCTGCCACCTTCCCGGATGAGCCCAACTTACAAAGATCGAGACCTCGCGTTCCG 840
Db 781 GTGTCTGCCACCTTCCCGGATGAGCCCAACTTACAAAGATCGAGACCTCGCGTTCCG 840
QY 841 CCCACAACCTACATTTGGGCACTGACTCAGAGCTGGCGATAGGGACACACGTTCTAG 900
Db 841 CCCACAACCTACATTTGGGCACTGACTCAGAGCTGGCGATAGGGACACACGTTCTAG 900
QY 901 GCCCGAGCGCCCTGTGCGCTGTGGGAGCTGGGAAGCCCGGAGGGGCTCCAGCGGG 960
Db 901 GCCCGAGCGCCCTGTGCGCTGTGGGAGCTGGGAAGCCCGGAGGGGCTCCAGCGGG 960
QY 961 ACTGGGGCTCTATCTACTCCCGAGTTTCCCAAGCTGGTACGCTGAGCCGCCACAGCCTCAT 1020
Db 961 ACTGGGGCTCTATCTACTCCCGAGTTTCCCAAGCTGGTACGCTGAGCCGCCACAGCCTCAT 1020
QY 1021 TGGAGAGTTCCTGCGCTGCAAGTGCAGCTCCCGATCCCTGCTGCTCCCGGGCACCC 1080
Db 1021 TGGAGAGTTCCTGCGCTGCAAGTGCAGCTCCCGATCCCTGCTGCTCCCGGGCACCC 1080
QY 1081 TGTGTCTCTCAGACTTCTTGTGAAGGGCCCAACAGGCGCTGGGCGTGGCGCTGGCAG 1140
Db 1081 TGTGTCTCTCAGACTTCTTGTGAAGGGCCCAACAGGCGCTGGGCGTGGCGCTGGCAG 1140
QY 1141 AAAGGAGGAGTCTCAGAGTGTCTGAATGGAAGTGTGGAGGCACTCGACATCTCCG 1200
Db 1141 AAAGGAGGAGTCTCAGAGTGTCTGAATGGAAGTGTGGAGGCACTCGACATCTCCG 1200
QY 1201 CCCTTCTGCTTCTATCTAGTACGCTCCCTGATTTAACAGAGATTCCGACAGTTCCTTGT 1260
Db 1201 CCCTTCTGCTTCTATCTAGTACGCTCCCTGATTTAACAGAGATTCCGACAGTTCCTTGT 1260
QY 1261 GCTGTGCTGCAACAAGGACATTTGAGGCTGATCTCTCTTAACCTCCTCCTAGTGTGGCC 1320
Db 1261 GCTGTGCTGCAACAAGGACATTTGAGGCTGATCTCTCTTAACCTCCTCCTAGTGTGGCC 1320
QY 1321 ACCTCAAACTCCCGTCCCAAGCAGAGGAGCGGTAGCACTAAATGTTGGAGACTCCG 1380
Db 1321 ACCTCAAACTCCCGTCCCAAGCAGAGGAGCGGTAGCACTAAATGTTGGAGACTCCG 1380
QY 1381 ATACTTCCTGCTGACTCCCGCTCTTTCAATCTCGGGGCTCCCAACACCGCTTCTCC 1440
Db 1381 ATACTTCCTGCTGACTCCCGCTCTTTCAATCTCGGGGCTCCCAACACCGCTTCTCC 1440

QY 1441 AGAGTGACCTAATCCAGTGT 1460
Db 1441 AGAGTGACCTAATCCAGTGT 1460

RESULT 2
RNRELAXT
LOCUS R.norvegicus mRNA for transcriptional regulator, Relax. 1491 bp mRNA linear ROD 06-MAY-1997
DEFINITION R.norvegicus mRNA for transcriptional regulator, Relax.
ACCESSION Y10619
VERSION Y10619.1 GI:2072737
KEYWORDS Relax; transcriptional regulator.
SOURCE Rattus norvegicus.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 1491)
AUTHORS Ravassard,P., Chetail,F., Mallet,J. and Icard-Liepkalos,C.
TITLE Relax, a novel rat BHLH transcriptional regulator transiently
expressed in the ventricular proliferating zone of the developing
central nervous system
JOURNAL J. Neurosci. Res. 48 (2), 146-158 (1997)
MEDLINE 97276390
PUBMED 9130143
REFERENCE 2 (bases 1 to 1491)
AUTHORS Ravassard,P.
TITLE Direct Submission
JOURNAL Submitted (20-JAN-1997) P. Ravassard, CNRS UMR 9923, Bat. CERV1,
Hopital de la Pitie Salpetriere, 83 Bd. de l'Hopital, F-75013
Paris, FRANCE
FEATURES
source
1..1491
/organism="Rattus norvegicus"
/strain="Wistar"
/db_xref="taxon:10116"
/clone_lib="RT-PCR"
459..1103
/function="transcriptional regulator"
/codon_start=1
/product="Relax"
/protein_id="CAA71630.1"
/db_xref="GI:2072738"
/db_xref="SPTREMBL:O08718"
/translation="MAPHPLDAPTIQVQETQOFPFGASDHEVLSSNTPSPPTLVPR
DCSEAAGDCRGTSKLRARGGRRKPKSELASKRRRRKANDRRNRHNLNSA
LDLUNGVLPTFPDDAKLTKIETLFAHNYIWAQTLLRIADHSFYGPPEPPVPCGELGS
PGGSSGDWGSIIYSPVQAGSLSPASLEFFPGLQVSPSPSCLPGLTVFSDFL"
BASE COUNT 307 a 487 c 413 g 284 t
ORIGIN
Query Match 100.0%; Score 1460; DB 10; Length 1491;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1460; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GCAGGTAGCGAGGAGGAGTCCCTGGGCGCCCGTGTGATTGGCCGTCGACAGGCA 60
Db 1 GCAGGTAGCGAGGAGGAGTCCCTGGGCGCCCGTGTGATTGGCCGTCGACAGGCA 60
QY 61 GCAGCCCGGCGAGGACGCTCTCTGGTCCGGGCGAGAGCAGATAAGCGTCCAGGGGACACA 120
Db 61 GCAGCCCGGCGAGGACGCTCTCTGGTCCGGGCGAGAGCAGATAAGCGTCCAGGGGACACA 120
QY 121 CGATTAGAGCTCAGAAGTCCCTCTGGGTCTCACCACCTGCACAGAGCGCGAGACCCCT 180
Db 121 CGATTAGAGCTCAGAAGTCCCTCTGGGTCTCACCACCTGCACAGAGCGCGAGACCCCT 180
QY 181 CCGAGCTTCTTGTGCTCCAGAGCAATTTACTCCAGGCGAGGCGCTCGAGCTCAG 240
Db 181 CCGAGCTTCTTGTGCTCCAGAGCAATTTACTCCAGGCGAGGCGCTCGAGCTCAG 240
QY 241 CAAAACCTCGAAGGAGGAGGAGGTTACGCTATCCACCGCTGCTTACTGCTGACACCC 300
```

```
|||||
Db 241 CAAAACCTCGAAGCAGCAGAGGGTTCAGCTATCCACCGCTGCTGACTCTGACCAACC 300
QY 301 GCAGCTCTCTGTTCTTTTGGAGCCCGAGTAAGTAACTTTAGGAACCTTCCAAAGG 360
Db 301 GCAGCTCTCTGTTCTTTTGGAGCCCGAGTAAGTAACTTTAGGAACCTTCCAAAGG 360
QY 361 TAGAAGAGGGAGTGGTGGGCTACTCTAGTCCCGGTGGAGTGAACCTTAAGTCAGAG 420
Db 361 TAGAAGAGGGAGTGGTGGGCTACTCTAGTCCCGGTGGAGTGAACCTTAAGTCAGAG 420
QY 421 ACTGTACACACCCCTCTTCATTTTCCCAACCTCAGATGCGCTCATCCCTTGGATG 480
Db 421 ACTGTACACACCCCTCTTCATTTTCCCAACCTCAGATGCGCTCATCCCTTGGATG 480
QY 481 CCCCCACCATCCAAAGTGTCCCAAGAGACCCAGCAACCCCTTCCCGAGCCTCGGACCAG 540
Db 481 CCCCCACCATCCAAAGTGTCCCAAGAGACCCAGCAACCCCTTCCCGAGCCTCGGACCAG 540
QY 541 AAGTGTCTAGTTCCAAATTCACCCACCTAGCCCACTCTGTACCGAGGGACTGCTCG 600
Db 541 AAGTGTCTAGTTCCAAATTCACCCACCTAGCCCACTCTGTACCGAGGGACTGCTCG 600
QY 601 AAGCAGAAGCAGTGAAGTGGGAGTGGGAGTGGGAGTGGGAGTGGGAGTGGGAGTGG 660
Db 601 AAGCAGAAGCAGTGAAGTGGGAGTGGGAGTGGGAGTGGGAGTGGGAGTGGGAGTGG 660
QY 661 GCAACAGGCCCCAAGAGCGAGTTGGCACTGAGCAAGCAGCAGCAAGCCCGCGCAAGAAG 720
Db 661 GCAACAGGCCCCAAGAGCGAGTTGGCACTGAGCAAGCAGCAGCAAGCCCGCGCAAGAAG 720
QY 721 CCACAGCCGGAGCGCAACCGCATGCACACCTTAACCTCCGCTGGATGCGCTGGCG 780
Db 721 CCACAGCCGGAGCGCAACCGCATGCACACCTTAACCTCCGCTGGATGCGCTGGCG 780
QY 781 GTGTCTGCCCCACTTCCCGGATGAGCCAACTTACAAAGATCGAGACCCCTCGCTTCG 840
Db 781 GTGTCTGCCCCACTTCCCGGATGAGCCAACTTACAAAGATCGAGACCCCTCGCTTCG 840
QY 841 CCCCAACTACATTTGGGCACTGACTCAGAGCTGGCGATAGCGGACCAAGCTTCTACG 900
Db 841 CCCCAACTACATTTGGGCACTGACTCAGAGCTGGCGATAGCGGACCAAGCTTCTACG 900
QY 901 GCCCCAGGCCCTGTGCTGCTGGGAGCTGGGAGCCCGGGAGGGGCTCCAGCGCG 960
Db 901 GCCCCAGGCCCTGTGCTGCTGGGAGCTGGGAGCCCGGGAGGGGCTCCAGCGCG 960
QY 961 ACTGGGGCTCTATCTACTCCCCAGTTTCCCAAGCTGGTAGCTGAGCCCCACAGCCTCAT 1020
Db 961 ACTGGGGCTCTATCTACTCCCCAGTTTCCCAAGCTGGTAGCTGAGCCCCACAGCCTCAT 1020
QY 1021 TGGAGGAGTTCCCTGGCCTGAGTGGCCCGAGTCCCATCCTGTCTCTCCCGGGCACCC 1080
Db 1021 TGGAGGAGTTCCCTGGCCTGAGTGGCCCGAGTCCCATCCTGTCTCTCCCGGGCACCC 1080
QY 1081 TGGTGTCTCAGACTTCTTGTGAAGGCCCAACAGGCCCTGGCGGTGGCGCTGGCAG 1140
Db 1081 TGGTGTCTCAGACTTCTTGTGAAGGCCCAACAGGCCCTGGCGGTGGCGCTGGCAG 1140
QY 1141 AAGGGAGGGAGTCAGAGCTGCTGAATGGAAGTGTGGAGGACTCGAGCATCTCGC 1200
Db 1141 AAGGGAGGGAGTCAGAGCTGCTGAATGGAAGTGTGGAGGACTCGAGCATCTCGC 1200
QY 1201 CCTTCTGGCTTTCATTTAGTCAGTCCCTGATTTAACAGGATTCGCACAGTTCCTTGCT 1260
Db 1201 CCTTCTGGCTTTCATTTAGTCAGTCCCTGATTTAACAGGATTCGCACAGTTCCTTGCT 1260
QY 1261 GCTGTGCTGCACAAAGGACATTCAGGCTGATCTCCTTTAACCCCTCCTCAGTGTGGCC 1320
Db 1261 GCTGTGCTGCACAAAGGACATTCAGGCTGATCTCCTTTAACCCCTCCTCAGTGTGGCC 1320
QY 1321 ACCTCAAACTCCCGCTCCCAAGCAGGAGAGCCGTAGCACTAAATAGTTGGGAGACTCCC 1380
|||||
```

Search completed: April 8, 2003, 17:16:12
Job time : 4065 secs

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 8, 2003, 05:37:59 ; Search time 373 Seconds
(without alignments)
8814.792 Million cell updates/sec

Title: US-09-595-947C-1

Perfect score: 1460
Sequence: 1 gcaggtacgagagagcag.....agagtacctaataccagtgt 1460

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 1

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 80%
Maximum Match 100%
Listing first 45 summaries

- Database : N_Geneseq_101002:*
- 1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:*
 - 2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:*
 - 3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:*
 - 4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:*
 - 5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:*
 - 6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:*
 - 7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:*
 - 8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:*
 - 9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:*
 - 10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:*
 - 11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:*
 - 12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:*
 - 13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:*
 - 14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:*
 - 15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:*
 - 16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:*
 - 17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:*
 - 18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:*
 - 19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:*
 - 20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:*
 - 21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:*
 - 22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:*
 - 23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:*
 - 24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	1460	100.0	1491	19	AAV42512

ALIGNMENTS

RESULT 1
AAV42512
ID AAV42512 standard; cDNA; 1491 BP.

XX AAV42512;
AC
XX
DI 05-OCT-1998 (first entry)
XX
DE CDNA encoding a novel BHLH protein designated RELAX.
XX
KW Basic helix-loop-helix; BHLH; RELAX; Rat Embryonic Longitudinal Axis;
KW control; gene expression; transcriptional activator; targeting;
KW protein expression; central nervous system; CNS; treatment;
KW nervous system disorder; ss.
XX
OS Rattus sp.
XX
FH Key Location/Qualifiers
FT 459..1103
CDS /*tag= a
FT /product= RELAX
FT
XX
PN W09827206-A2.
XX
PD 25-JUN-1998.
XX
XX 19-DEC-1997; 97WO-FR02368.
PF
XX 19-DEC-1996; 96FR-0015651.
PR
XX (RHON) RHONE-POULENC RORER SA.
PA
PI Mallet J, Ravassard P, Icard-Liepkalns C;
XX
XX WPI; 1998-362775/31.
DR P-PSDB; AAW62991.
XX
XX Basic helix-loop-helix polypeptide and related nucleic acid - with
PT transcriptional activity, for targeting expression of genes to
PT central nervous system and treatment of nervous disease
XX
PS Claim 6; Page 20; 28pp; French.
XX
XX The present sequence encodes a basic helix-loop-helix (BHLH) type
CC protein, designated RELAX (Rat Embryonic Longitudinal Axis) protein.
CC The protein is used to control and participate in gene expression,
CC by acting as transcriptional activator, strictly dependent on the
CC presence of an intact E box (CANNTG), particularly for targeting
CC expression of proteins to the central nervous system (CNS). The
CC nucleic acid sequence can be used to treat nervous system disorders,
CC and antisense sequences can be used to control mRNA transcription.
XX
XX Sequence 1491 BP; 307 A; 487 C; 413 G; 284 T; 0 other;
Query Match 100.0%; Score 1460; DB 19; Length 1491;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1460; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GCAGTAGCAGAGAGAGAGTCCCTGGGCCCCCGCTGCTGATGGCCGTCGACAGGCA 60
Db 1 GCAGTAGCAGAGAGAGAGTCCCTGGGCCCCCGCTGCTGATGGCCGTCGACAGGCA 60
Qy 61 GCAGCCCCGAGGAGGAGGCTCTGCTGGGCGAGAGAGAGTAAAGCGTGCCAGGGGACACA 120
Db 61 GCAGCCCCGAGGAGGAGGCTCTGCTGGGCGAGAGAGAGTAAAGCGTGCCAGGGGACACA 120
Qy 121 CGATTAGCAGCTCAGAAGTCCCTCTGGGTCTCCACCTGCACAGAGGCGGAGGCCCT 180
Db 121 CGATTAGCAGCTCAGAAGTCCCTCTGGGTCTCCACCTGCACAGAGGCGGAGGCCCT 180
Qy 181 CCGAGCTTCTTTGCTGCCCTCCAGAGCAATTTACTCCAGGCGGAGGCCCTCGACTCAG 240
Db 181 CCGAGCTTCTTTGCTGCCCTCCAGAGCAATTTACTCCAGGCGGAGGCCCTCGACTCAG 240
Qy 241 CAAACTTCGAAGCGAGAGAGGGTTGAGCTATCCACCGCTGCTGACTGTGACACCC 300
Db 241 CAAACTTCGAAGCGAGAGAGGGTTGAGCTATCCACCGCTGCTGACTGTGACACCC 300

Db 241 CAAAACCTTCGAAGCAGCAGAGGGGTTTCAGCTATCCACCGCTGCTGACTGTGACCTGACCAACC 300
QY 301 GCAGCTCTCTGTCTTTTGGAGCCCGAGTAAGTAACTATAGTAACATTTAGGAACCTCCAAAGGG 360
Db 301 GCAGCTCTCTGTCTTTTGGAGCCCGAGTAAGTAACTATAGTAACATTTAGGAACCTCCAAAGGG 360
QY 361 TAGAAGAGGGAGTGGGTGGCGTACTCTAGTCCCGGTGGAGTACCTCTAAATCAGAG 420
Db 361 TAGAAGAGGGAGTGGGTGGCGTACTCTAGTCCCGGTGGAGTACCTCTAAATCAGAG 420
QY 421 ACTGTACACCCCTTCCATTTTTTCCCAACCTCAGATGGCGCTCATCCCTTGGATG 480
Db 421 ACTGTACACCCCTTCCATTTTTTCCCAACCTCAGATGGCGCTCATCCCTTGGATG 480
QY 481 CGCCACCATCCAGTGTCCCAAGAGACCAGCAACCTTTCCCGGAGCCTCGGACCAG 540
Db 481 CGCCACCATCCAGTGTCCCAAGAGACCAGCAACCTTTCCCGGAGCCTCGGACCAG 540
QY 541 AAGTGTCTAGTTCATTTCCACCCCTAGCCCTAGCCCACTCTCTACGAGGAGTGTCTCG 600
Db 541 AAGTGTCTAGTTCATTTCCACCCCTAGCCCTAGCCCACTCTCTACGAGGAGTGTCTCG 600
QY 601 AAGCAGAGCAGTGTACTGCCGAGGACATCGAGGAGTCTCGTGGCGCGGAGGGC 660
Db 601 AAGCAGAGCAGTGTACTGCCGAGGACATCGAGGAGTCTCGTGGCGCGGAGGGC 660
QY 661 GCAACAGGCCCAAGAGCGAGTTGGCACTGAGCAAGCAGCAGCAAGCCGCGCAAGAAG 720
Db 661 GCAACAGGCCCAAGAGCGAGTTGGCACTGAGCAAGCAGCAGCAAGCCGCGCAAGAAG 720
QY 721 CCAACAGCCGGAGCGCAACCCGATGCGCAACCTTAACCTCCGCGTGGATGCGCTCGCG 780
Db 721 CCAACAGCCGGAGCGCAACCCGATGCGCAACCTTAACCTCCGCGTGGATGCGCTCGCG 780
QY 781 GTGTCTGCCACCTTCCGAGTACGCGCAAACTTACAAAGATCGAGACCTTGCCTTCG 840
Db 781 GTGTCTGCCACCTTCCGAGTACGCGCAAACTTACAAAGATCGAGACCTTGCCTTCG 840
QY 841 CCCACAATACATTTGGGCACTGACTCAGACGCTCGCATAGCGACCAACAGCTTCTACG 900
Db 841 CCCACAATACATTTGGGCACTGACTCAGACGCTCGCATAGCGACCAACAGCTTCTACG 900
QY 901 GCCCGAGCCCTGTGCCCTGTGGGAGCTGGGAAGCCCGGAGGGGCTCCAGCGCG 960
Db 901 GCCCGAGCCCTGTGCCCTGTGGGAGCTGGGAAGCCCGGAGGGGCTCCAGCGCG 960
QY 961 ACTGGGCTCTATCTACTCCCAAGTTCCCAAGCTGTAGCTGAGCCCAACAGCTCAT 1020
Db 961 ACTGGGCTCTATCTACTCCCAAGTTCCCAAGCTGTAGCTGAGCCCAACAGCTCAT 1020
QY 1021 TGGAGGAGTTCCTGGCTGAGTGCAGCTCCCATCCTGTCTCTCCCGGCAACC 1080
Db 1021 TGGAGGAGTTCCTGGCTGAGTGCAGCTCCCATCCTGTCTCTCCCGGCAACC 1080
QY 1081 TGGTGTCTCAGACTCTTGTGAAGGCCCAACAGCCCTGGCGGTGGCGCTGGCAG 1140
Db 1081 TGGTGTCTCAGACTCTTGTGAAGGCCCAACAGCCCTGGCGGTGGCGCTGGCAG 1140
QY 1141 AAAGGGAGGAGTCAAGCTGTCTGAAATGGAAGTGTAGTGGAGCACTCGAGCATCTCG 1200
Db 1141 AAAGGGAGGAGTCAAGCTGTCTGAAATGGAAGTGTAGTGGAGCACTCGAGCATCTCG 1200
QY 1201 CCTTCTGGCTTTCATAGTCAAGTCTGATTTAACAGGATTCGACAGTTCTCTTGT 1260
Db 1201 CCTTCTGGCTTTCATAGTCAAGTCTGATTTAACAGGATTCGACAGTTCTCTTGT 1260
QY 1261 GCTGTGCTGCACAAAGGACATTCAGGCTGATCTCTTAACCTCTCTAGTGTGGCC 1320
Db 1261 GCTGTGCTGCACAAAGGACATTCAGGCTGATCTCTTAACCTCTCTAGTGTGGCC 1320
QY 1321 ACCTCAAACTCCCGCTCCCAAGCAGAGGAGCGGTAGCACTAAATAGTTGGGAGCTCCC 1380
Db 1321 ACCTCAAACTCCCGCTCCCAAGCAGAGGAGCGGTAGCACTAAATAGTTGGGAGCTCCC 1380

QY 1381 ATACTTCTGTGACTCGGCCCTCTTTCAAATCTGCGGCCTCCAAACACCGCTTTCTCC 1440
Db 1381 ATACTTCTGTGACTCGGCCCTCTTTCAAATCTGCGGCCTCCAAACACCGCTTTCTCC 1440
QY 1441 AGAGTGACCTAATCCAGTGT 1460
Db 1441 AGAGTGACCTAATCCAGTGT 1460

Search completed: April 8, 2003, 11:37:01
Job time : 374 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 8, 2003, 11:23:35 ; Search time 122 Seconds
(without alignments)
10497.254 Million cell updates/sec

Title: US-09-595-947C-1
Perfect score: 1460
Sequence: 1 gcaggtagcagagagcag.....agagtgcctaatccagtg 1460

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 593429 seqs, 438583890 residues

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 80%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_NA:
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq.*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
---------------	-------	----------------	--------	----	----	-------------

No matches found

Search completed: April 8, 2003, 18:58:39
Job time : 124 secs

GenCore version 5.1.4_p5_4578
Copyright(c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 8, 2003, 06:37:40 ; Search time 2258 Seconds
(without alignments)
10471.838 Million cell updates/sec

Title: US-09-595-947C-1
Perfect score: 1460
Sequence: 1 gcaggtacgagagagcag.....agagtacctaataccagtg 1460

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 80%
Maximum Match 100%
Listing first 45 summaries

- Database : EST:*
- 1: em_estba:*
 - 2: em_esthum:*
 - 3: em_estin:*
 - 4: em_estmu:*
 - 5: em_estov:*
 - 6: em_estpl:*
 - 7: em_estro:*
 - 8: em_hic:*
 - 9: gb_est1:*
 - 10: gb_est2:*
 - 11: gb_hic:*
 - 12: gb_est3:*
 - 13: gb_est4:*
 - 14: gb_est5:*
 - 15: em_estfun:*
 - 16: em_estom:*
 - 17: gb_gss:*
 - 18: em_gss_hum:*
 - 19: em_gss_inv:*
 - 20: em_gss_pln:*
 - 21: em_gss_vrt:*
 - 22: em_gss_fun:*
 - 23: em_gss_mam:*
 - 24: em_gss_mus:*
 - 25: em_gss_other:*
 - 26: em_gss_pro:*
 - 27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description

No matches found

Search completed: April 8, 2003, 17:56:21
Job time : 2258 secs


```

/strain="C57BL/6J"
/db_xref="FANTOM.DB:2010001M19"
/db_xref="MGI:MGI:1907403"
/db_xref="taxon:10090"
/clone="2010001M19"
/sex="male"
/tissue_type="small intestine"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
1..1540
/gene="Atoh5"
241..885
/gene="Atoh5"
/notes="atonal homolog 5 (Drosophila)"
data source:MGI, source key:MGI:893591, evidence:ISS
putative"
/codon_start=1
/protein_id="BAB25411.1"
/db_xref="GI:12841942"
/db_xref="MGI:MGI:893591"
/translation="MAPHPDALTIOVSPETQOPFCASDHEVLSNSTPSSHLLIPR
DCSEAVGDCRGSRKLRRRGGRNRPKSELAKRRSRKXANDRRNRNMNLISA
LDALRGVLPFPDALKTVETLRFNAHYIWAQTIRIADHSLYGPEPVPVCGELGS
PGGSGNDGSIYSPVSGAGNLSPTASLEEFGLGVPSYLLPGALVFSDFL"
BASE COUNT 306 a 485 c 397 g 352 t
ORIGIN
Query Match 4.1%; Score 60; DB 11; Length 1540;
Best Local Similarity 100.0%; Pred. No. 4.8e-20;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 762 GCGTGGATGCGTGGCGGTGTCCTGCCACCTCCCGGATGAGCGCAACTACAAAG 821
|||||
DB 544 GCGTGGATGCGTGGCGGTGTCCTGCCACCTCCCGGATGAGCGCAACTACAAAG 603
|||||

RESULT 3
BG808248 600 bp mRNA linear EST 20-DEC-2001
LOCUS 2083-52 Mouse E14.5 retina lambda ZAP II Library Mus musculus cDNA,
DEFINITION mRNA sequence.
ACCESSION BG808248
VERSION BG808248.1 GI:17955225
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Mu.X., Zhao,S., Pershad,R., Hsieh,T.-F., Scarpa,A., Wang,S.W.,
White,R.A., Beremand,P.D., Thomas,T.L., Gan,L. and Klein,W.H.
Gene expression in the developing mouse retina by EST sequencing
and microarray analysis
Nucleic Acids Res. 29 (24), 4983-4993 (2001)
21671825
Contact: Klein WH
Department of Biochemistry and Molecular Biology
University of Texas M.D. Anderson Cancer Center
Box 117, 1515 Holcombe Blvd., Houston, TX 77030, USA
Tel: 713 792 3646
Fax: 713 790 0329.
Location/Qualifiers
1..600
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone_lib="Mouse E14.5 retina lambda ZAP II Library"
/tissue_type="neural retina"
/dev_stage="embryonic day 14.5 post-fertilization"
/notes="Vector: pAMP10 (Gibco); Cloned unidirectionally.
Primer: Oligo dt. RNA isolation: cytoplasmic RNA preps
(Mannatis); Cloning technique: CUA Cloning (Clontech,
Life Technologies); Average insert size: 1.8 Kb;
Insertion site: TACGTCCACTGAATTCGTAGTG--->. Other

information regarding entire library may be found at
http://pga.swmed.edu/Data/Libraries/microarray_cdna_library_ies.htm."
BASE COUNT 94 a 238 c 161 g 107 t
ORIGIN
Query Match 1.6%; Score 23; DB 12; Length 600;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 732 GAGCGCAACCGCATGCACACCT 754
|||||
DB 4 GAGCGCAACCGCATGCACACCT 26
|||||

RESULT 4
AL540071 804 bp mRNA linear EST 16-FEB-2001
LOCUS AL540071 LTI_FL013_FBrnl Homo sapiens cDNA clone CS0DF035YD19 5
DEFINITION prime, mRNA sequence.
ACCESSION AL540071
VERSION AL540071.1 GI:12869886
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 804)
Li,W.B., Gruber,C., Jesse,J. and Polayes,D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
Location/Qualifiers
1..804
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DF035YD19"
/clone_lib="LTI_FL013_FBrnl"
/dev_stage="pooled tissue from post conception fetuses (20
week, 24 week and 26 week)"
/lab_host="DH10B"
/notes="Organ: Fetal brain; Vector: pCMVSPORT 6; 1st strand
cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-stranded cDNA was digested with Not I
and cloned into the Not I and Eco RV sites of the
pCMVSPORT 6 vector. Library was constructed by Life
Technologies. Contact : Feng Liang Life Technologies, a
division of Invitrogen 9800 Medical Center Drive Rockville
, Maryland 20850, USA Fax : (1) 301 610 8371 Email :
fliang@lifetech.com URL :
http://fulllength.invitrogen.com"
BASE COUNT 154 a 270 c 278 g 99 t
ORIGIN
Query Match 1.6%; Score 23; DB 9; Length 804;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 831 CTGCGCTTCGCCACACTACAT 853
|||||
DB 770 CTGCGCTTCGCCACACTACAT 792
|||||

RESULT 5
BQ178789 814 bp mRNA linear EST 30-APR-2002
LOCUS BQ178789
DEFINITION UI-M-EV0-bwt-k-07-0-UI.r1 NIH-BMAP-EV0 Mus musculus cDNA clone
IMAGE:5701758 5', mRNA sequence.
ACCESSION BQ178789
VERSION BQ178789.1 GI:20354281

```

KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
JOURNAL 1 (bases 1 to 814)
COMMENT NIH-MGC Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: Dr. James Lin, University of Iowa
CDNA Library preparation: Dr. M. Bento Soares, University of Iowa
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)

FEATURES
source
Seq primer: pYX-5.
1. .814
Location/Qualifiers
/organism="Mus musculus"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:5701758"
/clone_lib="NIH_BMAP_EV0"
/tissue_type="whole brain"
/dev_stage="embryo 15.5 dpc"
/lab_host="DH10B (T1 phage resistant)"
/note="Organ: brain; Vector: pYX-Asc; Site 1: EcoR I;
Site 2: Not I; The library was constructed according to
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured mRNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with an
oligo-dT primer containing a Not I site. Double stranded
cDNA was size selected according to mRNA size fraction,
ligated with EcoR I adaptor, digested with Not I, and then
cloned directionally into pYX-Asc vector. The library tag
sequence located between the Not I site and the polyA tail
, is GTGCGTGGAA. This library was created for the
University of Iowa Mouse Brain Molecular Anatomy Project
(BMAP). 'Gene Discovery in the Developing Mouse Nervous
System', supported by National Institutes of Mental Health
(NIMH), Hemin Chin, Ph.D., program coordinator."

BASE COUNT 151 a 268 c 291 g 102 t 2 others
ORIGIN

Query Match 1.6%; Score 23; DB 14; Length 814;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 732 GAGCGCAACCGCATGCACACCT 754
|||||
DB 663 GAGCGCAACCGCATGCACACCT 685
|||||

RESULT 6
BG854922 1269 bp mRNA linear EST 29-MAY-2001
LOCUS 1024041C06.y2 C. reinhardtii CC-1690, normalized, Lambda Zap II
DEFINITION Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION BG854922
VERSION BG854922.1 GI:14236106
KEYWORDS EST.
SOURCE Chlamydomonas reinhardtii.
ORGANISM Chlamydomonas reinhardtii.

REFERENCE
AUTHORS Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
1 (bases 1 to 1269)
Chlamydomonadaceae; Chlamydomonas.
Grossman, A., Davies, J., Federspiel, N., Harris, E., Lefebvre, P.,
McDermott, J.P., Silflow, C., Stern, D. and Surzycki, R.

TITLE Analyses of the Chlamydomonas reinhardtii Genome: A Model,
Unicellular System for Analyzing Gene Function and Regulation in
Vascular Plants; project phase 2
JOURNAL Unpublished (2000)
COMMENT Contact: Charles Hauser
DCMB Box 91000
Duke University
Durham, NC 27708-1000
Tel: 919 613 8159
Fax: 919 613 8177
Email: chauser@duke.edu.

FEATURES
source
Location/Qualifiers
1. .1269
/organism="Chlamydomonas reinhardtii"
/strain="CC-1690 wild type mt+ 21gr"
/db_xref="taxon:3055"
/clone_lib="C. reinhardtii CC-1690, normalized, Lambda Zap
II"
/note="Vector: pBluescript II SK-; Site 1: EcoRI; Site 2:
XhoI; This library, constructed by John Davies and Jeffrey
McDermott, combines cDNAs from CC-1690 cells grown to
mid-log phase in TAP (acetate-containing) medium in the
light, TAP medium in the dark, HS (minimal) medium in
ambient levels of CO2 and HS medium bubbled with 5% CO2.
PolyA mRNA was purified from each sample, pooled and cDNA
synthesized. The cDNA was directionally cloned into lambda
ZAP II (Stratagene) in the EcoRI (5') and XhoI (3') sites.
pBluescript II SK- plasmids were excised from the lambda
ZAP clones by superinfection with ExAssist (Stratagene)
phage. The library was normalized using method 4 described
in Bonaldo et al (1996) Genome Research 6: 791-806."

BASE COUNT 558 a 231 c 444 g 12 t 24 others
ORIGIN

Query Match 1.6%; Score 23; DB 12; Length 1269;
Best Local Similarity 100.0%; Pred. No. 1.7;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 653 CGGAGGGCGCACAGCCCAAGA 675
|||||
DB 1244 CGGAGGGCGCACAGCCCAAGA 1266
|||||

RESULT 7
BE936551 319 bp mRNA linear EST 02-OCT-2000
LOCUS RC5-NT0053-310800-024-E01 NT0053 Homo sapiens cDNA, mRNA sequence.
DEFINITION BE936551
ACCESSION BE936551
VERSION BE936551.1 GI:10462627
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 319)
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,
Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,
M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
Simpson, A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=6t2-rc5-NT0053-310
 800-024-E01st3-2000-08-31&t4=1)
 Seq primer: puc 18 forward
 High quality sequence stop: 319.

FEATURES

Source

1. .319
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="NT0053"
 /dev_stage="Adult"

/note="Organ: nervous_tumor; Vector: puc18; Site_1: SmaI;
 Site_2: SmaI; A mini-library was made by cloning products
 derived from ORESTES PCR (U.S. Letters Patent application
 No. 196,716 - Ludwig Institute for Cancer Research)
 profiles into the pUC 18 vector. Reverse transcription of
 tissue mRNA and cDNA amplification were performed under
 low stringency conditions."

BASE COUNT

53 a 87 c 106 g 73 t

ORIGIN

Query Match 1.4%; Score 21; DB 12; Length 319;
 Best Local Similarity 100.0%; Pred. NO. 15;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1112 AACAGGCCCTGGCGGTGGGC 1132

|||||

Db 280 AACAGGCCCTGGCGGTGGGC 300

RESULT 8

AQ753599

LOCUS

HS_2117_A2_H03_MR_CIT Approved Human Genomic Sperm Library D Homo

sapiens genomic clone Plate=2117 Col=6 Row=O, DNA sequence.

DEFINITION

ACCESSION AQ753599

VERSION AQ753599.1

KEYWORDS GSS.

SOURCE human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 464)

Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,

Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and

Hood,L.

Sequence-tagged connectors: A sequence approach to mapping and

scanning the human genome

Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)

99380589

Contact: Mahairas GG, Wallace JC, Hood L

High Throughput Sequencing Center

University of Washington

401 Queen Anne Avenue North, Seattle, WA 98109, USA

Tel: (206) 616-3618

Fax: (206) 616-3887

Email: jwallace@u.washington.edu

Clones may be purchased from Research Genetics (infotresgen.com).

BAC end Web Server: http://www.htsc.washington.edu

Plate: 2117 row: 0 column: 6

Seq primer: M13 Reverse

Class: BAC ends

High quality sequence stop: 464.

Location/Qualifiers

1. .464

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="plate=2117 Col=6 Row=O"

/clone_lib="CIT Approved Human Genomic Sperm Library D"

/sex="male"

/note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in

E-Coli DH10B"

FEATURES

Source

1. .464

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="plate=2117 Col=6 Row=O"

/clone_lib="CIT Approved Human Genomic Sperm Library D"

/sex="male"

/note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in

E-Coli DH10B"

BASE COUNT

125 a 114 c 101 g 123 t 1 others

ORIGIN

Query Match 1.4%; Score 21; DB 17; Length 464;
 Best Local Similarity 100.0%; Pred. NO. 16;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 492 CAAGTGTCCCAAGACCCAG 512

|||||

Db 106 CAAGTGTCCCAAGACCCAG 126

RESULT 9

AZ506899

LOCUS

1M0348G13F Mouse 10kb plasmid UUGC1M library Mus musculus genomic

clone UUGC1M0348G13 F, DNA sequence.

DEFINITION

ACCESSION AZ506899

VERSION AZ506899.1

KEYWORDS GSS.

SOURCE house mouse.

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

1 (bases 1 to 536)

Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,

Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly

,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.

and Wright,D., Weiss,R.

Mouse whole genome scaffolding with paired end reads from 10kb

plasmid inserts

Unpublished (2000)

Contact: Robert B. Weiss

University of Utah Genome Center

University of Utah

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT

84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: ddunne@genetics.utah.edu

Insert Length: 10000

Std Error: 0.00

Plate: 0348

row: G

column: 13

Seq primer: CGTGTAAACGACGCGCCAGT

Class: plasmid ends

High quality sequence stop: 536.

Location/Qualifiers

1. .536

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone_lib="UUGC1M0348G13"

/clone_lib="Mouse 10kb plasmid UUGC1M library"

/sex="Male"

/lab_host="E. Coli strain XL10-Gold, Tl-resistant, P-"

/note="vector: PWD42nv; Purified genomic DNA from M.

musculus C57BL/6J (male) was obtained from the Jackson

Laboratory Mouse DNA Resource

(http://www.jax.org/resources/documents/dnares/). The DNA

was hydrodynamically sheared by repeated passage through a

0.005 inch orifice at constant velocity. The sheared DNA

was blunt end-repaired with T4 DNA polymerase and T4

polynucleotide kinase. Adaptor oligonucleotides were

ligated to the blunt ends in high molar excess. The

adapted DNA was purified and size-selected for a

10.5 kb range using preparative agarose

electrophoresis. Vector DNA was prepa

of PWD42 (g14732114|gb|AF129072.1)

with adaptors complementary to the

purified. The sheared, adapted mo

adapted vector DNA, and transforme

chemically-competent E. coli XL10-Go

and selected for ampicillin resistance;

```

BASE COUNT      156 a      119 c      81 g      180 t
ORIGIN

Query Match      1.4%; Score 21; DB 17; Length 536;
Best Local Similarity 100.0%; Pred. NO. 17;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 916 TGCCTGTGGGAGCTGGGAA 936
|||||
Db 406 TGCCTGTGGGAGCTGGGAA 426

RESULT 10
BE263765      718 bp      mRNA      linear      EST 13-JUL-2000
LOCUS 6011941122F1 NIH_MGC_7 Homo sapiens cDNA clone IMAGE:3537940 5',
DEFINITION mRNA sequence.
ACCESSION BE263765
VERSION BE263765.1 GI:9137311
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 718)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Plate: LNCM222 row: 1 column: 05
High quality sequence stop: 649.
Location/Qualifiers
1..718
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NIH_MGC_7"
/tissue_type="small cell carcinoma"
/cell_line="MGC3"
/lab_host="DH10B (phage-resistant)"
/Note="Organ: lung; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5',
adaptor: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
```

BASE COUNT 115 a 195 c 260 g 148 t
ORIGIN

Query Match 1.4%; Score 21; DB 10; Length 718;
Best Local Similarity 100.0%; Pred. NO. 18;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1112 AACAGGCCCTGGCGGTGGGC 1132
|||||
Db 536 AACAGGCCCTGGCGGTGGGC 556

RESULT 11
BI910102 829 bp mRNA linear EST 16-OCT-2001
LOCUS 603067946F1 NIH_MGC_118 Homo sapiens cDNA clone IMAGE:5217154 5',
DEFINITION mRNA sequence.
ACCESSION BI910102
VERSION BI910102.1 GI:16173443
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 829)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11545 row: k column: 11
High quality sequence stop: 828.
Location/Qualifiers
1..829
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NIH_MGC_118"
/tissue_type="leukocyte"
/lab_host="DH10B"
/Note="Vector: pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV
(destroyed); RNA source leukocytes from anonymous pool of
non-activated adult donors. Library is oligo-dT primed
and directionally cloned (EcoRV site is destroyed upon
cloning). Average insert size 1.7 kb, insert size range
1.2-3.3 kb. Library is normalized and enriched for
full-length clones and was constructed by C. Gruber
(Invitrogen). Research Genetics tracking code 027. Note:
this is a NIH_MGC Library."

BASE COUNT 146 a 224 c 293 g 166 t
ORIGIN

Query Match 1.4%; Score 21; DB 13; Length 829;
Best Local Similarity 100.0%; Pred. NO. 18;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1112 AACAGGCCCTGGCGGTGGGC 1132
|||||
Db 550 AACAGGCCCTGGCGGTGGGC 570

RESULT 12
BF204175 913 bp mRNA linear EST 06-NOV-2000
LOCUS 601867625F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:4110222 5',
DEFINITION mRNA sequence.
ACCESSION BF204175
VERSION BF204175.1 GI:11097761
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 913)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LNCM996 row: a column: 07
High quality sequence stop: 714.
Location/Qualifiers
1..913
/organism="Homo sapiens"